

**IMT School for Advanced Studies Lucca**

Lucca, Italy

**Collective Behaviour in Digital Societies**

PhD in Institutions, Markets and Technologies -  
Curriculum in Economics, Management, and Data Science

XXXII Cycle

**By**

**Abhishek Samantray**

**2020**



**The dissertation of Abhishek Samantray is approved.**

Program Coordinator: Prof. Massimo Riccaboni, IMT School for Advanced Studies Lucca, Italy

Supervisor: Prof. Massimo Riccaboni, IMT School for Advanced Studies Lucca, Italy

Co-Supervisor: Prof. Paolo Pin, University of Siena, Italy

The dissertation of Abhishek Samantray has been reviewed by:

Prof. Arun Gautham Chandrasekhar, Stanford University, USA

Prof. Giorgio Fagiolo, Scuola Superiore Sant'Anna, Italy

**IMT School for Advanced Studies Lucca**

**2020**



To Ranjita



# Contents

<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xii</b>
<b>Acknowledgements</b>	<b>xiv</b>
<b>Vita and Publications</b>	<b>xv</b>
<b>Abstract</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Minds & Collective Behaviour . . . . .	2
1.1.1 Information, Intellect & Decisions . . . . .	3
1.1.2 Role of Social Networks . . . . .	3
1.1.3 Creator of Complexity: Kahneman or Tirole? . . . .	4
1.1.4 Importance in Society, Environment & Economy . .	5
1.1.5 Various Perspectives for Analysing Behaviour . . .	6
1.2 Technology & Digital Platforms . . . . .	7
1.2.1 Digital Economies . . . . .	7
1.2.2 Digital Societies . . . . .	8
1.3 Research Questions, Analysis of Effects & Mechanisms . .	10
1.3.1 Are digital learners influenced by their peers? . . .	10
1.3.2 When does homophily increase polarization? . . . .	12
1.3.3 Does politicization of news affect participation? . .	13
1.4 Implications of Research Findings . . . . .	14

<b>2</b>	<b>Peer Influence in Scratch, an Educational Platform</b>	<b>16</b>
2.1	Introduction . . . . .	16
2.2	Data: Scratch Community . . . . .	21
2.2.1	Users, Projects . . . . .	24
2.2.2	Measures used for analysis . . . . .	25
2.2.3	Assortativity in Behaviour . . . . .	26
2.3	Peer Influence Analysis . . . . .	28
2.3.1	Methods . . . . .	28
2.3.2	Results . . . . .	34
2.4	Mechanism of Peer Influence . . . . .	44
2.5	Discussion . . . . .	50
2.5.1	Limitations stemming from data . . . . .	51
2.5.2	Validity and interpretation of results . . . . .	53
2.5.3	Behaviours analysed in this study . . . . .	56
2.5.4	Some topics for further investigation . . . . .	58
<b>3</b>	<b>Polarization in Twitter, a Social Media Platform</b>	<b>59</b>
3.1	Introduction . . . . .	59
3.2	Methods . . . . .	62
3.2.1	Data . . . . .	62
3.2.2	Measuring Sentiment & Opinion . . . . .	62
3.2.3	Measure of Homophily . . . . .	63
3.2.4	Measure of Polarization . . . . .	64
3.3	Results . . . . .	66
3.3.1	Negative effect of Homophily on Polarization . . . . .	66
3.3.2	Joint Effect of Homophily and Credibility . . . . .	68
3.4	Discussion . . . . .	71
<b>4</b>	<b>Politicization in Guardian, a News Media Outlet</b>	<b>76</b>
4.1	Introduction . . . . .	76
4.2	Data . . . . .	80
4.3	Politicization of climate change . . . . .	81
4.3.1	Macro attributes . . . . .	85
4.3.2	Micro attributes . . . . .	87
4.4	Consumption perspective: Impact on collective discussion . . . . .	91



4.4.1	Impact of macro attributes . . . . .	95
4.4.2	Impact of micro attributes . . . . .	98
4.4.3	Mechanisms of politicized discussions . . . . .	100
4.5	Production perspective: Risk preference of authors . . . . .	107
4.6	Discussion . . . . .	111
<b>5</b>	<b>Conclusion</b>	<b>113</b>
5.1	Summary of Results . . . . .	113
5.2	Data is King & Context is Queen. They Dance Together! . .	115
5.3	Looking Ahead . . . . .	116
<b>A</b>	<b>Peer Influence in Scratch</b>	<b>118</b>
A.1	Producer Types . . . . .	118
A.2	Consumption Behaviour . . . . .	120
A.2.1	Consumption Baskets . . . . .	121
A.2.2	Consumption Specificity . . . . .	123
A.3	Tables & Figures . . . . .	124
<b>B</b>	<b>Polarization in Twitter</b>	<b>130</b>
B.1	Cointegration Test . . . . .	130
B.2	Estimation using Vector Error Correction Model . . . . .	132
B.2.1	Granger-Causality Tests . . . . .	135
B.3	A Model of Polarization in Social Networks . . . . .	137
<b>C</b>	<b>Politicization in Guardian</b>	<b>143</b>
C.1	Tables & Figures . . . . .	143
C.1.1	Joint effects of macro and micro attributes . . . . .	143
C.1.2	Impact of macro attributes . . . . .	148
C.1.3	Impact of micro attributes . . . . .	153
C.1.4	Mechanisms of politicized discussions . . . . .	158
	<b>References</b>	<b>164</b>

# List of Figures

1	PEER INFLUENCE MECHANISM (PRODUCTION POPULAR- ITY) . . . . .	18
2	PRODUCTION AND CONSUMPTION PERSPECTIVES OF A PROJECT IN SCRATCH 2.0 . . . . .	21
3	JOINING OF USERS AND CREATION OF PROJECTS . . . . .	24
4	ASSORTATIVE MIXING IN FRIENDSHIP NETWORK . . . . .	27
5	ACCUMULATION OF LOVE-ITS . . . . .	33
6	PEER INFLUENCE ONE MONTH AHEAD . . . . .	35
7	ROBUSTNESS CHECK FOR PEER INFLUENCE ONE MONTH AHEAD . . . . .	36
8	PERSISTENCE OF PEER INFLUENCE . . . . .	38
9	PEER INFLUENCE CHANNEL FOR PRODUCTION POPULAR- ITY: CREATION OF PROJECTS . . . . .	45
10	ROBUSTNESS CHECK FOR PEER INFLUENCE CHANNEL . . .	48
11	PEER INFLUENCE CHANNEL IS VALID IN GENERAL . . . . .	49
12	POLARIZATION OF BELIEFS, HOMOPHILY IN COMMUNI- CATION . . . . .	67
13	BELIEF UPDATING MODEL . . . . .	70
14	COLLECTIVE DISCUSSION . . . . .	79
15	SAMPLE DISCUSSION . . . . .	82
16	POLITICIZATION FROM POLITICAL INCLINATIONS OF EN- TITIES . . . . .	90

17	MECHANISMS FOR COLLECTIVE ATTENTION & DISCUSSION	102
18	RISK AVERSION COEFFICIENTS: QUARTERLY ESTIMATES .	108
19	RISK AVERSION COEFFICIENTS: MONTHLY ESTIMATES . .	109
20	TYPES OF PRODUCERS . . . . .	118
21	CONSUMPTION COMMUNITIES . . . . .	121
22	POLARIZED CONSUMPTION: EVIDENCE OF CONSUMPTION SPECIFICITY . . . . .	122
23	SCRATCH PLATFORM . . . . .	124
24	SAMPLE SENSITIVITY TEST VISUALIZATION: PART 1 . . . .	162
25	SAMPLE SENSITIVITY TEST VISUALIZATION: PART 2 . . . .	163

# List of Tables

1	HETEROGENOUS INFLUENCE (SUSCEPTIBILITY) . . . . .	40
2	GRANGER-CAUSALITY TESTS . . . . .	68
3	SECTION SPLITS FOR CLIMATE CHANGE RELATED ARTI- CLES . . . . .	81
4	CATEGORIES FOR ARTICLES . . . . .	86
5	TYPES OF NAMED ENTITIES . . . . .	89
6	AVERAGE TREATMENT EFFECTS OF MACRO ATTRIBUTE .	97
7	AVERAGE TREATMENT EFFECTS OF MICRO ATTRIBUTES .	99
8	VARIABLES DESCRIPTION . . . . .	125
9	VARIABLES DESCRIPTION (..CONTINUED..) . . . . .	126
10	MODEL VARIABLES, CONFOUNDERS . . . . .	127
11	BALANCE OF COVARIATES . . . . .	128
12	PEER/NETWORK EFFECT . . . . .	129
13	NAIVE LOOK AT RELATIONSHIPS IN TIME SERIES . . . . .	130
14	ADF TEST STATISTIC ESTIMATES . . . . .	131
15	JOHANSEN COINTEGRATION TEST . . . . .	133
16	ESTIMATES OF VEC MODELS . . . . .	134
17	HYPOTHESIS TESTS ON COINTEGRATING VECTOR $\beta$ . . . . .	135
18	EFFECT ON DISCUSSION SIZE . . . . .	144
19	EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS .	145

20	EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS	
	146	
21	EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION . . . .	147
22	BALANCE OF COVARIATES, SAMPLE SIZE, AUTHORS RE- TAINED . . . . .	148
23	EFFECT ON DISCUSSION SIZE . . . . .	149
24	EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS .	150
25	EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS	151
26	EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION . . . .	152
27	BREAKING CORRELATIONS WITH CONFOUNDERS . . . . .	153
28	EFFECT ON DISCUSSION SIZE . . . . .	154
29	EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS .	155
30	EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS	
	156	
31	EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION . . . .	157
32	POLITICIZED COLLECTIVE DISCUSSIONS: MECHANISM 1	159
33	POLITICIZED COLLECTIVE DISCUSSIONS: MECHANISM 2	160

## Acknowledgements

I offer my deep gratitude to Professor Riccaboni and Professor Pin for their consistent support on various activities and decisions during the course of my doctoral studies. Several aspects of my research work has benefited from the time they have spent with me, as supervisors and co-authors, during various stages of progress. I thank all faculty members for their several helpful remarks and discussions. I thank Professor Chandrasekhar and Professor Fagiolo for their time to review this thesis. I thank Professor Valentina Bosetti for providing a visiting period at Bocconi University. I thank Dr. Christos Nicolaides for being a co-author in the last chapter of the thesis.

I thank all staff members for facilitating official paper work very efficiently. I acknowledge generous financial support received from the school on various occasions.

I thank all my colleagues and friends for their kindness, discussions, and humor. I will cherish all the football matches we watched, all ping pong matches we played, the food we cooked and dined, and very nice outings and trips. The time I stayed at Lucca for my doctoral work is a memorable one.

I thank my parents and family for their love and support. I acknowledge the opportunity to visit Southern India during one of the summer vacations.

In this year, 2020, we are seeing the unfolding of the pandemic. My prayers lie with all people who have experienced any consequences closely, and a sincere thanks to many officials and people who have been helping to fight this.

# Vita

## EDUCATION

- 2016-2020      PhD in Economics, Management and Data Science,  
IMT School for Advanced Studies Lucca, Italy
- 2014-2016      Master in Economic Theory and Econometrics,  
Toulouse School of Economics, France
- 2009-2014      Master of Science (Integrated Prog.) in Mathematics,  
National Institute of Science Education and Research,  
India

## GRANTS

- 2019-2020      Frontier Proposal Fellowship, IMT School
- 2016-2019      IMT PhD Scholarship & Travel Grants
- 2015-2016      TSE Master Scholarship
- 2009-2014      INSPIRE Scholarship, Dept. of Science & Technology,  
Govt. of India

## EVENTS

- PhD Workshop 2019 Scuola Superiore Sant'Anna, Pisa
- 4th European Conference on Social Networks (2019), Zurich
- Tut. on Text Analysis, Int. Conf. on Comp. Soc. Sc. 2019, Amsterdam
- Tut. on Social Media Analysis, Complexity72H 2019 Workshop, Lucca
- Tut. on Network Analysis, Complex Net. Conf. 2018, Cambridge (UK)
- Long. Network Analysis, 47th GESIS Spring Seminar 2018, Cologne
- Int. Conference on Synthetic Populations (2017), Lucca
- 5th TSE Student Workshop (2015), Toulouse
- Financial Econometrics Conference (2015), Toulouse
- NPDE-TCA Advanced Level Training Program (IIT-B), Mumbai

## Publications

1. Samantray A., Riccaboni M. (2019) "Peer Influence in Large Dynamic Network: Quasi-experimental Evidence from Scratch." In: Aiello L., Cherifi C., Cherifi H., Lambiotte R., Lió P., Rocha L. (eds) *Complex Networks and Their Applications VII*. COMPLEX NETWORKS 2018. Studies in Computational Intelligence, vol 813. Springer, Cham.  
doi: [https://doi.org/10.1007/978-3-030-05414-4\\_24](https://doi.org/10.1007/978-3-030-05414-4_24)
2. Samantray, A., Pin, P. (2019) "Credibility of climate change denial in social media." *Palgrave Communications (Nature Research)* 5, 127.  
doi: <https://doi.org/10.1057/s41599-019-0344-4>
3. Samantray A., Riccaboni M. (2020) "Peer influence of production and consumption behaviour in an online social network of collective learning." *Online Social Networks and Media* 18, 100088.  
doi: <https://doi.org/10.1016/j.osnem.2020.100088>
4. Samantray A., Nicolaidis C. (2020) "Collective attention to politicization of information: The case of news articles on climate change" *Working Paper*



# Presentations

## CONFERENCES (2017-2020)

1. "Peer Influence in Large Dynamic Network: Quasi-experimental Evidence from Scratch." *7th International Conference on Complex Networks and their Applications (Complex Networks 2018)*, Cambridge (UK). Session: Social Networks
2. "Peer influence of production and consumption behaviour in an online social network of collective learning." *5th International Conference on Computational Social Science (IC2S2 2019)*, Amsterdam. Track: Social Influence
3. "Polarization of beliefs due to homophily in communication and credibility of fake news." *4th European Conference on Social Networks (EUSN 2019)*, Zurich. Session: Communication Networks

## WORKSHOPS & SEMINARS (2017-2020)

1. Seminar for PhD students (Apr 2018), IMT 2nd Research Symposium, (Nov 2018), AXES Lab meeting (Nov 2018), *Internal Seminars at IMT School*, Italy
2. "Peer influence of production and consumption behaviour in an online social network of collective learning." *PhD Workshop 2019, Scuola Superiore Sant'Anna*, Pisa. Discussant: Giorgio Fagiolo
3. "Credibility of climate change denial in social media." *La Strada Seminar (Dec 2019)*, Bocconi University, Milan.

# Abstract

The number of social and economic activities on digital platforms has been increasing since the last two decades, and especially in the last decade. Several such platforms also provide opportunities for interactions among participating users, either as a part of the primary activities on such platforms or as a secondary feature. Such interactions form the basis of the emergence of collective behaviour on individual platforms, wherein users' interactions affect the aggregate state of the platform and at the same time the aggregate state of platform feeds into how users interact among themselves. Digital platforms turn into digital societies by providing many ways to conduct users' interactions driven activities, that were traditionally conducted in the physical world, in efficient and scalable ways. Apparently, these two societies – digital and real world – exist simultaneously and have feedback effects in shaping outcomes in both the societies on various factors including behaviour, beliefs, opinions, and others. This thesis contains three essays including (1) peer influence on creation of new projects in the learning environment Scratch, (2) polarization of climate change beliefs due to homophily in interactions on the social media Twitter, and (3) effects on collective attention due to political framing of climate change by the news media Guardian. Particular emphasis is laid on statistical inference of the effects and hypothesizing mechanisms behinds such effects.

The production and consumption of projects in the Scratch community, a digital platform developed by MIT Media Lab where users, usually young children, learn to program by creating and sharing projects, is analysed using the data for the

first five years after its launch. In particular, investigation is done to discover if users are influenced by the popularity of their peers' projects and their peers' preferences for consuming specific baskets of projects. The major challenges in this type of analysis is to provide parsimonious models for complexities of interactions on the platform and to disentangle peer influence from homophily in the vast network of behaviours and friendships. Homophily is a term widely used in social networks studies to describe friendship or tie formations that arise due to similarities in behaviours or common attributes between participating agents in the formation of such ties. The analysis reveals that while Scratchers' consumption preference is not influenced by their peers, the popularity of their projects is significantly influenced by their peers in short and long terms. A large proportion of the influence from peers is mediated via Scratchers' creation of new projects, which highlights Scratchers' subsequent decisions in response to existing popularity of peers' projects. These insights can potentially help in incorporation of behaviour-driven designs in future educational technologies.

Producer-consumer business models is at the heart of several social networking sites. Activities on such sites range from meeting friends, exchanging messages, propagating messages, advertisements, and others. Lack of regulations on information posting and limitations of computer-assisted information checks therefore provide opportunities for people's beliefs to be polarized due to the spread of fake information in such social networks. Homophily in communication creates groups of people or agents with bounded beliefs about the reality, and hence can polarize a society. Such homophily in digital media has been termed as echo chambers which intuitively promotes the notion that people hear nothing more than what they already believe. Using evidence from 11 years of Twitter conversations on the climate change topic, an em-

pirical analysis is conducted on the effect of homophily in communication patterns on the polarization of beliefs about the reality of climate change. The analysis reveals a counter-intuitive result that increasing levels of homophily in communication predicts decreasing levels of polarization in beliefs in the long run. To understand better the mechanism of the effect of homophily on polarization, a model is developed that shows how polarization can emerge due to the joint effects of precision of misinformation propagating in a social network and homophily in communication among agents in the social network with differing beliefs. Credibility of fake news, modelled as precision of misinformation, circulating in the social network can lead to acceptance of the fake news (depending on agents' susceptibility to it), thereby changing beliefs and creating polarization. The model shows that fake news can not polarize the society unless it has a minimal level of credibility, irrespective of the level of homophily in communication patterns. This throws a light on perhaps the most intuitive but usually the forgotten factor of information – credibility. While the results show that the climate change sceptic exchanges of messages on the social media Twitter do not carry enough credibility to create large scale polarization in society, they also provide useful indications to directly or indirectly quantify the emergence of credibility of information in digital platforms, and also to shift attention of technology from detecting fake stories to detecting fake stories which the society might find them to be credible.

Digital news platforms have become outlets for engaging discussions. They provide journalists and publishers with several dimensions to gauge the acceptance of their articles and at the same time provides readers with simple tools to participate in discussions. It has been long acknowledged that news media frame their articles. As an instance of political framing, to understand how the politicization climate change

articles influences the collective attention and discussion on such articles, articles published in the Guardian until 2018 are used to examine whether and how politicization influences readers' collective discussion. The results suggest that (unknown) factors of perception associated to an article when it is categorized in 'Politics' section positively impacts the collective attention and engagement received on articles. Estimates also suggest that mentions of political inclinations of the entities within an article impact such collective attention. In particular, a large proportion of such participation is found to be mediated by discussions becoming politicized by past contextual entities related to the article but strictly absent in the article, thereby suggesting a temporal effect of perception. In addition, a large proportion of the impact of political mentions on users' engagement is mediated by users who join discussion being influenced by past politically oriented contextual entities. Although no evidence is found to support that authors or journalists might be enjoying increasing marginal benefits from collective attention to their articles that result from their choices about mentions of political entities within the main texts of climate change articles, the results highlight that their choices do impact the readers' perceptions and participations (and potentially the intensity of climate change action in real world).

Overall, this series of research has contributed to improving how collective behaviour is shaped in digital societies in relation to peer influence in educational media, polarization of beliefs in social media, and political framing of articles in news media. Hopefully, these results would be of interest to general audience and researchers in fields of social sciences, economics, marketing, media and communications, and applied data science.

# Chapter 1

## Introduction

Collective behaviour in a given society is an emergent process or a global pattern which exists as a result of individual activities in the society, but may not be fully explained without taking into account the interactions among individuals, the motives of participation, and the intelligence/perception that emerges or is shaped within each individual due to such interactions. In this thesis we will see three studies on collective behaviour with an aim to investigate some of the important contemporary issues on three different digital platforms.

(1) Whether people are aware about it or not, social influence exists in varying intensities across various interactive platforms. The intensities can be explained on a range between random clicks to clicks with known causes and consequences. Social influence on digital platforms forms that dimension of collective behaviour where users' cognitive states are affected by the feedbacks from the state of affairs of the society and induce changes in how users behave on the platforms.

(2) Beliefs form the units of blocks based on which people decide whether to act and how to act. In the digital age, there are several formal and informal sources of information. Social media platforms have become a place for the spread of news and information, whose validity is subject to various factors. Fake information not only creates wrong beliefs in individuals but also across the population which can make each

individual be convinced about his or her belief just because everybody believes it. While the fake notion of information may not be exclusive to digital platforms, the notion that you know what others might be believing as well creates a common society-wide perception. These abilities, before the age of online social networks, were concentrated with mass media outlets in forms of newspapers and news on the television. However the digital generation has empowered every individual and organization with equal abilities to influence mass perceptions. Hence, propagation of misinformation is an important issue in the digital era.

(3) Public opinion on important issues like politics is largely shaped by news media. Public opinions in turn shape social and economic policies for the future democracy. Mass media communications therefore play a big role in the feedback between democratic policies and public opinion. Digital news media create a different environment for readers when compared to newspaper readership or social media. Digital media put published articles' reputation in the hands of the public to assess the reliability of the content and also to drive further discussion.

## **1.1 Minds & Collective Behaviour**

Collective behaviour refers to the observed behavioural patterns of a group that emerges due to the ways individuals or agents in the group acquire and process information, interact and communicate with other agents in the group, and reasons or motivations behind their actions. An important and distinct aspect of collective behaviour is that it can not be understood by studying patterns of individual behaviours alone, and feedbacks exist between individual agents and group-level outcomes.

Such global outcomes that emerge from complex interactions among individual agents of the society may or may not be intuitive, with traceable or untraceable mechanisms. The term 'agent' has a meaning with respect to the units of participants in a given society, which is not limited just to human societies. Collective behaviour in several circumstances consists of simultaneous feedbacks between the global state of the given society and the states of its participating units. Collective behaviour can

result in consequences undesirable for the society in spite of seemingly harmless activities performed by individuals. It can be the other way around as well: collective outcomes desirable by a society can emerge from individual actions which were not guided by a desire for such outcomes.

### **1.1.1 Information, Intellect & Decisions**

Every action by a social agent is an outcome of information accumulated from various sources, the ways in which the agent processes such information to form beliefs, and deciding to act based on the final beliefs. Collective behaviour is an aggregate manifestation of heterogeneous atomistic information, belief formations, and decisions of individuals with varying abilities. Each of these components at individual levels can vary enormously especially when individual social agents are human beings unlike other animals. For instance it is rare to see human dynamics similar to a flock of birds flying together, and bird dynamics similar to a mob protesting for a purpose. Comparing humans with birds shows a clear pattern that these two social agents have extremely varying abilities in information acquisition, processing, and decision making. While these abilities seem to be almost homogeneous across all birds of a given species, the same can not be said about human beings due to their differences in cognitive capabilities and their differences in control over handling the mental faculties. Cognitive actions can occur at multiple levels composed of agents with varying identities and outcomes at each level can receive or give feedbacks to other levels. In this thesis, we do model (or describe) information sources, belief updates, and decision-making of agents wherever possible in the course of investigating research questions.

### **1.1.2 Role of Social Networks**

Connectedness is an inherent characteristic of social agents, and this is perhaps what makes each agent a social agent. There different kinds of interactions that individuals can have between each other – friendships,



communications, mentoring, etc. – with different intensities and durations. A social network, conceptually, describes the composition of social agents with heterogeneous characteristics or behaviours and interactions among them. Interactions in social networks form the substrate on which information, opinions or beliefs, and decisions about behaviours or adoptions spread between agents. Such interactions result in various outcomes or processes such as copying, learning, innovating to improve, innovating to differentiate, and others. The aggregate outcomes of a group, conceived as a social network, therefore not only depend on behaviours of individual agents, but also on the nature and dynamics of interactions among the agents. In this thesis, we do model the various interactions among agents as required by the nature of research question to be studied.

### **1.1.3 Creator of Complexity: Kahneman or Tirole?**

In this section, description is made about the importance of motivation behind human actions, and how differences in motivations can result in emergence of different collective behaviours and social or economic outcomes.

Economy is a method of efficiently organizing the aspects of social beings, and particularly in case of human beings, which are related to sustained survival with limited resources. Several traditional assumptions to model economy usually incorporate maximizing self-interests or monetary profits as the primary objective of actions made by individuals and firms (composed of groups of individuals and resources). At the same time, individuals also engage in helping others and firms also engage in non-monetary motives like building open-source software.

The complexities of social interactions among different kinds of agents (individuals, groups, firms, governments, etc.) and organizational structures (e.g., economic markets, cultures, religions, institutions, nations, etc.) are largely motivated due to heterogeneous reasons. Broadly, such motivations can be categorized as competition (as a way of trying to be exclusive) and co-operation (as a way of trying to be inclusive). While

firms in the same economic market may compete for profits and co-operate with other markets for developing supply chains, individuals identified with one culture or religion may co-operate among themselves but compete or exclude individuals with different identities. At the end of the day, the collective behaviour that emerges in different conceptual layers of societies is essentially a combination of social and economic activities, with consequences on the environment which sustains these activities. This shows that human beings have heterogeneous motivations for various activities and this must be understood as a part of the observed emergence of collective behaviour. However irrational and rational behaviours of human beings and firms (containing human beings) can be meaningful only if they are described conditional on certain pre-conceived assumptions about their expected optimal behaviour in various situations. Let us look at an example. Two birds released from the same cage flew in east and west directions, one in each direction. Which one do you think is rational? In this thesis, we do not model motivations or optimal actions of interacting agents in a collective environment. The results can be considered naive observations by looking at the data. (Behavioural assumptions, wherever made, have been kept to minimum.)

### **1.1.4 Importance in Society, Environment & Economy**

Collective behaviour is observed in different circumstances of grouping like crowds, social movements, and other voluntary activities. Understanding collective behaviour can provide insights into initiating and organizing required changes in society, especially the unfavourable aspects of aggregate outcomes. It can also provide insights into prescribing norms for individual behaviour during emergence of spontaneous situations like panics, mass hysteria, clashes between social groups with differing ideologies, and others. Modeling collective behavior has the potential to deliver effective and efficient ways to predict and control collective outcomes. Understanding trends of collective behaviour in social media platforms can provide insights into various matters that need prescriptions, including emergency situations like management activi-

ties during natural disasters.

Social and economic activities affect the environment. While driving a new car does not seem to create global warming, the same action when done by many people deciding individually for themselves can have large explanatory power to explain rising temperatures in cities. In matters that concern the environment, the public understanding may not develop until there is a real experience. Public understanding is confounded by several factors – fake information, lack of perception of the role of individual activities, self-reinforcing and biased beliefs, incentives of government agents and institutions in the society, and others. Regulating industries and imposing taxes on firms may create perpetual loopholes if the collective demand for sustainable products and services is not in place. On the other hand, creating social awareness and behavioural changes can actually have long lasting impacts both from demand and supply perspectives.

In economic and corporate management activities, there are increasing adoptions of business models and developments of algorithms that can harness collective behaviour. Facebook, for instance, has a business model that revolves around insights from collective and interactions, activities, and interests. Algorithms for recommender systems, which have been shown to be of value at Amazon, are examples that use digital trails of collective behaviour. Product developments in firms happen due to co-operative interactions among employees. Linux is well known to have features that emerged beyond what it was originally intended for. A market emerges due to collective behaviour comprising competitive interactions among suppliers of innovations and budget constrained choosers of friend-recommended or expert-recommended innovations. Collective behaviour in form of co-operation emerges to create efficient, valuable, and robust supply chains.

### **1.1.5 Various Perspectives for Analysing Behaviour**

There has been an upsurge of data and methods to understand collective behaviour, due to increase in capacity to store and process big data.

Various disciplines including social sciences and humanities, sociology, economics, marketing, mathematics, computer science, physics, psychology, and others have contributed to the understanding of collective behaviour. Current research on various issues and methods are largely organized in computational social science which shares similarities with research in digital societies, behavioural economics, processes on and of networks, agent based models, social computing, complex networks, and cyber-emotions. New types of data are being analyzed which include phone traces, social networks data, web blogs and video, sensing of face-to-face interactions, crowdsourcing, and others. Methods for analysis of networks and text data emerged as the major contributors in making such analysis possible.

## **1.2 Technology & Digital Platforms**

Digital avenues have generated increasing attention for various activities. Several platforms provide services that include mass participation.

### **1.2.1 Digital Economies**

Although this thesis does not analyse digital economies, its narration will probably provide a better context to understand the information value contained in digital societies, which is discussed in the next section.

With rise in information and communication technology, several goods and services are traded on digital platforms. Examples include e-commerce, music download, software, etc. Most businesses from the early years were platform services. They were incorporated with several features to help customers choose informed products. Most of these features mimicked sellers trying to compete for visibility and reputation of their products within the platform. Platforms like eBay and Amazon attracted have attracted increasing traffic. Very soon, customers were able to get involved in the platform apart from transactions due to facilities for likes and reviews. Next product recommendations and personalized experiences became popular tools on platforms that attract huge traffic. Digital

platforms have also become hot-spots for marketing activities compared to traditional methods due to ease of use. Fintech has produced various innovative products providing standalone services like currency exchanges or integrated services like retail payments. An increasing number of startups in recent years are seeking customer attention through user experience designs and marketing efforts.

However, the overall trend in these types of platforms have minimal notions of societies because they tend to be largely transactional. Also facilities like review systems where large number of users can leave reviews does not really translate into a concept of societies because a particular user does not have incentives for repeating such activities once the product is bought. In a physical shop, this translates to leaving behind feedbacks, which do not incorporate active and sustained interactions among customers.

### **1.2.2 Digital Societies**

Digital societies are for-profit or non-profit digital platforms, usually with large number of participating users, whose services provide social experiences to the users rather than benefits of only economic transactions as in digital economies. Examples of such platforms with pre-defined set of activities include those like Wikipedia, Github, LinkedIn, news media and opinion blogs, while general social platforms for a range of activities include Youtube, Facebook and Twitter. It is very clear that the examples of digital platforms mentioned above are much different in nature from various digital economies mentioned in previous section. Clearly, these platforms provide social value beyond economic transactions.

What factors turn these platforms into societies? First, repeated and active interactions among participants is the key element in all the above examples of digital societies. Second, most of such platforms are based on prosumer models where users both produce and consume content on such platforms. Examples of prosumer models with separable boundaries between producers and consumers would include platforms like online courses and news media outlets.

Digital societies offer various services and interactions ranging from non-profit oriented collaborative activities as in Wikipedia and Github to profit-oriented platforms as in Facebook and Reddit. Activities include social networking, information propagation, job searching, building software, information organization, discussions, and many others. Digital platforms are composed of users, users' interactions among themselves, and the primary activities for which they are on the platform. In some platforms, the later two may not be distinguishable. Users' behaviour on digital platforms are subject to similar judgements and issues as they face in real-world interactions. For example. when two people interact in a social gathering due to shared interests also interact in similar way when online. What differentiates users' experiences in digital societies from physical real-world societies include (1) scale and efficiency of participation, and (2) the emergence of collective behaviour. While the former is driven by innovative technology, the later is strictly a social dimension.

How can information in digital societies help to create better societies, or what can we learn by investigating such platforms? Digital economies exploit collective intelligence to design new products or improve existing products. As well-known examples, Google uses information about linkages among web pages in its page rank algorithm to build useful search engine, and Amazon uses big data on users' collective interests and buying patterns to drive growth in business by building recommendation engines. Such things make digital platforms unique in terms of data sources for harnessing collective intelligence, since such data may not be easily available in physical shops. In an analogous way, digital societies not only contain traces of economic preferences wherever possible, but also a rich and diverse set of information on other aspects of individual and collective behaviour. Analysing such behaviour can provide fundamental insights into how people behave in digital spaces and real world. This is because the fundamental nature of human and collective behaviour is reflected in digital spaces due to the scale of participation and complex interactions.

Above, there is no argument made to suggest that digital societies are right or wrong in any way. What is being said is that it reflects fundamen-

tal nature of human social behaviour which would otherwise be difficult to infer in real world due to lack of large scale data. So if there is certain section of society to spread misinformation and another section with certain characteristics is susceptible to it, we would expect such behaviour to exist in digital societies as well. In that sense, existence of online echo chambers reveals the nature of communications within communities of people with exposure to self reinforcing beliefs. Understanding such nature can help policy-makers develop better policies to build the society and economy in a sustainable way, and also to engage citizens actively in important issues.

Studying digital societies would therefore help to understand demographic preferences and behaviour at granular levels, to promote democracy and participation of citizens, to identify social agents and issues that threaten democracy, to make the traditional media more transparent, to develop public understanding on global issues and propose policy changes, to build better educational and social platforms in future, and to create opportunities for cultural integration and better integrations.

## **1.3 Research Questions, Analysis of Effects & Mechanisms**

### **1.3.1 Are digital learners influenced by their peers?**

We study peer influence of production and consumption of projects in the Scratch community, an online platform developed by MIT Media Lab and targeted for young children, where users collectively learn programming by creating and sharing projects. We investigate if Scratchers are influenced by the popularity of their peers' projects and their peers' preferences for consuming from specific baskets of projects.

To estimate peer influence of a behaviour,<sup>1</sup> we use exact matching

---

<sup>1</sup>This is in reference to the following published work:

Samantray A., Riccaboni M. (2019) "Peer Influence in Large Dynamic Network: Quasi-experimental Evidence from Scratch." In: Aiello L., Cherifi C., Cherifi H., Lambiotte R., Lió P., Rocha L. (eds) *Complex Networks and Their Applications VII*. COMPLEX NETWORKS 2018. Studies in Computational Intelligence, vol 813. Springer, Cham

strategy to justify a random assignment of the peers' behaviour across experimental and control groups such that Scratchers in the experimental group have peers with higher degree of the behaviour under study. Next, conditional on interactions on the platform upto a given time, we measure peer influence as the difference in Scratchers' future behavioural changes across the two groups. This method ensures that peer influence is captured after controlling for alternative mechanisms (like homophily) that may lead to observed behavioural clustering in the followers network.

We find<sup>2</sup> that the popularity of Scratchers' projects is significantly influenced by the production popularity of their peers. Testing for heterogeneity in influence, we find that Scratchers are not influenced by specific peers who might have highly popular projects, instead it seems that they are influenced by just the aggregate popularity of all peers. We find that Scratchers who have a minimum activity of one month on the platform are more susceptible to peer influence. Scratchers with high tendency to create projects by rebuilding on existing projects on the platform tend to have significant improvements in their future production popularity (due to influence from peers' production popularity) only in the short term and not in the long run. We also disentangle a self decision making mechanism from other mechanisms that might explain the channel of influence: we find that a significant proportion of the estimated influence from peers is mediated via Scratchers' decision to create new projects. This highlights Scratchers' subsequent behavioural decisions in response to existing popularity of peers' projects.

We find evidence of polarized consumption patterns on the platform, i.e., there are certain groups of projects (discovered in an unsupervised manner based on co-consumption patterns) for which Scratchers have high specificity. We do not make claims about how such groups form on the platform - for example, whether it is a conscious choice or is a result of the way the platform is organized. However, we find that such

---

<sup>2</sup>This is in reference to the following published work:  
Samantray A., Riccaboni M. (2020) "Peer influence of production and consumption behaviour in an online social network of collective learning." *Online Social Networks and Media* 18, 100088



polarization is not a consequence of Scratchers being influenced by their peers' consumption patterns.

### 1.3.2 When does homophily increase polarization?

We investigate how people's beliefs can be polarized due to the spread of fake information in social networks. In particular, we model how the emergence of polarization can occur due to (i) precision or credibility of fake information during its inception into the society, and (ii) homophily in communication among people with differing beliefs. Homophily in communication creates groups of people with bounded beliefs about reality, and hence can polarize a society. Credibility of fake news circulating in the social network can lead to acceptance of the news (depending on agents' susceptibility to it), thereby changing beliefs and creating polarization. Hence both these factors, homophily in communication patterns and credibility of information, are important determinants of polarization.

We conduct an empirical investigation<sup>3</sup> of the effect of homophily in communication patterns on the level of polarization. For this, we use evidence from 11 years of Twitter conversations on the climate change topic. There are two important findings. (i) Homophily and polarization are entangled in a long run equilibrium, i.e., they are co-integrated, and they both mean-revert to deviations away from equilibrium. (ii) Homophily negatively affects polarization (Granger-causality sense), and the effect holds only in the long run.

To understand better the mechanism behind the negative effect of homophily on polarization, a parsimonious mathematical model is developed that shows how polarization can emerge due to the joint effects of precision of misinformation propagating in a social network and homophily in communication among agents in the social network with differing beliefs. Credibility of information contained in a message is modelled as the precision (or inverse of variance) of the distribution of beliefs

---

<sup>3</sup>This is in reference to the following published work:  
Samantray, A., Pin, P. (2019) "Credibility of climate change denial in social media." *Palgrave Communications* 5, 127

from which the message can be considered as a random outcome. Essentially, a particular communicated message says something about its underlying belief and how precise the belief is. The model predicts that fake news can not polarize the society unless it has a minimal level of credibility, irrespective of the level of homophily.

Based on above analyses, we infer that anti-climate tweets do not carry enough credibility to polarize the society.

### **1.3.3 Does politicization of news affect participation?**

Do choices made by news media to politicize articles related to climate change influence the collective discussion on such articles? Using articles published in the Guardian until 2018, we examine whether and how politicization influences readers' collective discussion.

We find a positive impact of politicization on readers' participation and response to the articles. We analysed impacts originating from (i) macro attributes: (unknown) factors of perception associated to an article when it is categorized in 'Politics' section, and (ii) micro attributes of politicization: political inclinations of the entities within an article. We find that the positive impact would be significant, in a counter-factual scenario, if an article positioned in any section other than 'Politics' were to be repositioned in 'Politics' section.

The estimates suggest that micro attributes of an article affect the number of users in its discussion, in particular, with at least 65% of such participation being mediated by discussions becoming politicized by contextual entities related to the article but strictly absent in the article, thereby suggesting a temporal effect of perception. We also find that the impact of micro attributes on total comments, social feedbacks, and users' engagement in discussion mediates with a high proportion (of at least 40%) through users who join discussion being influenced by politically oriented contextual entities not mentioned in the article.

We investigate the risk preference of authors to politicize the content of climate change articles. We do not find evidence that authors might have increasing marginal benefits from collective attention to their arti-

cles that result from their choices about micro attributes of politicization.

## 1.4 Implications of Research Findings

Today various online educational platforms facilitate collective ways of learning. The literature on peer influence on educational outcomes presents mixed evidences, i.e., both positive and negative influences. Since learning via educational platforms is gaining increasing interest, it is important for future policy designs to know the real effect of peers' activities on the choices and educational outcomes of users. On the Scratch platform which is targeted for young children to learn programming, we find that users are influenced by the production popularity of their peers, but not by their peers' consumption patterns. We believe that understanding such behavioural nature would be very helpful to design platforms in future or improve technical components in existing platforms such that the collective educational outcome can be greatly enhanced.

Polarization of beliefs is increasing in modern societies due to spread of fake information in digital media, even on issues like climate change which have extensive scientific documentation. In social networks, polarization may be confused with homophily in communication (communication among people having same beliefs) because they are highly correlated: polarization can cause homophily because it is a source of differentiation, and homophily can cause polarization via echo chamber effect where individual beliefs get reinforced. In the Twittersphere of climate change conversations during 2007-2017, only homophily causes polarization and not vice-versa. Also, the effect of homophily on polarization is negative, which is surprising. We are able to explain this using a mechanism where the updating of beliefs accounts for the relative credibility of fake information. The results potentially indicate that although detecting fake news is important, detection and prevention of "credible fake news" can be more helpful for the society at large. Probably, credibility of information is one of the least quantitatively explored areas of how people process information in various places (digital and physical) and in various contexts (social, environmental, economic) despite its major

significance.

Politicization of climate change in news media has the potential to alter people's perception about the climate change issue and hence the gravity for climate action. How people perceive an issue can drive future policies. We believe climate change action by the public precedes correct and neutral judgement about the issue both individually and collectively. It is therefore important that journalists and authors of news articles on climate change be aware about the impact they have on their audience regarding an issue as sensitive as climate change. It is equally important for the public to be aware about the fact that other people with whom they might be interacting or discussing about climate change may carry political inclinations from the articles they might have learnt about on the web or elsewhere. The findings show that the public should engage in some forms of scientific reading in order to gauge whether the perceptions and beliefs formed by reading news articles are in line with scientific evidences, or other forms of expert evidences and suggestions.

## Chapter 2

# Peer Influence in Scratch, an Educational Platform

### 2.1 Introduction

Today<sup>1</sup> various online platforms facilitate learning by means of collective activities. Peers' behaviour can play an important role in various aspects of one's learning process [3]. Peers' behaviour can play an important role in various aspects of one's learning process [3]. Knowing the real impact of peers activities on the choices and performance outcomes of users in learning environments is therefore important for business and economic policy designers. For instance, knowledge of such behavioural nature would be useful to design better platforms where the collective educational outcome is maximized.

How co-learners influence educational and social outcomes has been studied extensively in physical contexts like schools and universities [4,

---

<sup>1</sup>This chapter is based on the following two published works:

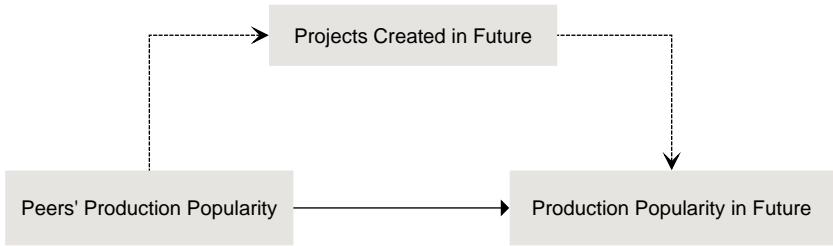
1. Samantray A., Riccaboni M. (2019) "Peer Influence in Large Dynamic Network: Quasi-experimental Evidence from Scratch." In: Aiello L., Cherifi C., Cherifi H., Lambiotte R., Lió P., Rocha L. (eds) *Complex Networks and Their Applications VII*. COMPLEX NETWORKS 2018. Studies in Computational Intelligence, vol 813. Springer, Cham [1],
2. Samantray A., Riccaboni M. (2020) "Peer influence of production and consumption behaviour in an online social network of collective learning." *Online Social Networks and Media* 18, 100088 [2].

5]. With the advent of various online education media and many users joining such sites, it is important to study peer influence in digital platforms as well. Such platforms usually encourage learning by various forms of collective interactions such as discussions in forums, building collaborative projects, private communications, and others [6]. A possibility of peer influence arises since users are usually aware of others' shared activities. In this study, we investigate peer influence in the Scratch platform which has a structure of learning through collective activities. Scratch, made public in 2007, is an online community designed by the MIT Media Lab for young people to learn programming. Scratchers produce and share visual projects built using programming codes. Scratchers consume others' projects in different ways which include viewing, commenting, loving, downloading, etc. Scratchers can know about the activities of their peers, other users whom they "follow" on the platform, via activity feeds and also by manual visits to their project pages. This creates a potential channel of influence on various behaviours. In the first five years of Scratch's public activities [7], during which about 1 million users joined the platform and about 2 million projects were created, we investigate peer influence on two behaviours – one relates to production of projects, and the other relates to consumption of projects. In particular we investigate (i) how the popularity of peers' projects influences Scratchers' future behaviour and production popularity, and (ii) how peers' consumption preference/specificity influences Scratchers' preferences to consume similar projects.

Understanding of peer influence estimation has been shaped by contributions from academics and practitioners in various fields including marketing, sociology, and economics. To infer peer influence, the most ideal situation would be to impute peers' behaviour at random and measure its average effect on Scratchers' behaviour. Marketing scientists have used such behavioural imputations in various online platforms to measure peer influence [8, 9]. However such experimental situations are usually not feasible, especially in non-artificial circumstances, for several reasons including ethics and permissions to perform such experiments [10, 11, 12]. In non-experimental settings, obtaining unbiased

estimates of peer influence is a challenging task because both individuals and their peers can affect each others' behaviour (reflection problem [13]), and so observed clustering of behaviour in networks is often a result of the following effects: own tendency for the behaviour, peers' influence on behaviour, and exogenous and endogenous network formation processes (homophily, selection, reciprocity, etc.) that lead to observed peers' behaviour. Dynamic observations help to separate changes in individual behaviour due to peers' influence from effects arising due to alternative mechanisms. Sociologists have used agent-based models [14, 15, 16, 17, 18] to explain the coevolution of network and behaviour. However, the state-of-the-art implementation of this method is computationally inefficient for the case when network is populated by a large number of agents. Economists have used estimation strategies that usually require strong assumptions [19, 20] and are specific to the structural models employed [21]. Sometimes exogenous component of peer influence (arising from past) is not estimated separately from the contemporaneous effect arising due to simultaneous determination of network formation and behavioural influence [22].

**Figure 1: PEER INFLUENCE MECHANISM (PRODUCTION POPULARITY)**



Popularity of peers' projects affects the popularity of Scratchers' projects in future (solid path). A significant proportion of such peer influence on production popularity is mediated via Scratchers' creation of new projects in future (dotted path).

We employ a quasi-experimental method to identify peer influence. We assume Scratchers have a Markov nature of decision making, i.e., their decisions about future actions (e.g., whom to follow, what to produce and consume) are based solely on the current state of activities on

the platform. Conditional on all activities upto a given time  $t$ , we estimate peer influence at  $t$  on a future time  $t + j$  as the effect of peers' behavioural state at  $t$  on Scratchers' subsequent change in behaviour upto  $t + j$ . The quasi-experiment consists of observations at two time periods –  $t$ ,  $t + j$  – and treatment status is assigned at  $t$  based on the intensity of peers' behaviour (high or low) at  $t^2$ . The treated group has Scratchers whose peers have high degree of behaviour under study. To conceptualize the treatment status as a random assignment, the control group is adjusted by matching exactly on personal and peers' characteristics of Scratchers in the treated group such that all confounding factors are balanced across the two groups. Below are the main results and contributions of our investigation:

- If peers' projects are popular (as measured by accumulated 'loves' on the projects) at  $t$ , the popularity of Scratchers' projects increases in future periods. This effect is persistent, and the marginal effect tends to decrease over time. We provide robustness checks for this finding to ensure that the estimates are not sensitive to certain threshold values chosen for the analysis.
- We find several evidences of susceptibility of Scratchers to peer influence. For example, a minimum engagement of about a month on the platform makes Scratchers more susceptible to such influence, however higher engagement does not necessarily increase the susceptibility.
- Remixing is a key property of the Scratch platform – users can build projects on top of existing projects by modifying or introducing new elements. Developers, users whose projects tend to be mostly remixed and not new projects, are highly susceptible to peers' performance in the short term. Free-style producers, Scratchers who create new and remixed projects in similar proportions, tend to be influenced in later periods only and not in the immediate period.

---

<sup>2</sup>The definition of treatment follows directly from the Markovian nature of decision making by Scratchers. The treatment, peers' behavioural state at  $t$ , is a measure that summarizes peers' behaviour upto  $t$ . It captures only the cumulative information of peers' behaviour upto  $t$  and neglects the historical pattern of its evolution.



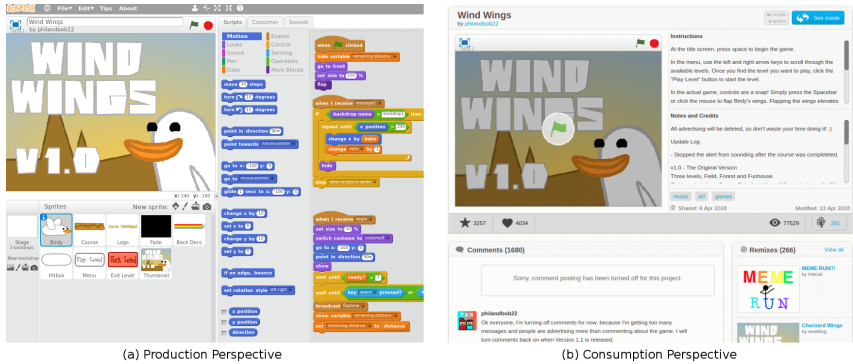
- As documented in Fig. 1, we investigate if the observed peer influence in production popularity is caused due to production-related decisions made by Scratchers. A large fraction of the total effect of peers’ production quality on Scratchers’ future production popularity is mediated via their creation of new projects in future. This channel emphasizes the role of decision-making under influence of peers’ behaviour. We provide details of using causal mediation analysis [23] to obtain this finding. We conduct robustness checks to ensure validity of the finding.
- Aggregate consumption patterns of Scratchers is highly polarized. They tend to consume projects from specific communities only. The ‘specific communities’ used are identified by us using unsupervised techniques, and signify ‘consumption baskets’.<sup>3</sup> However, we find that consumption polarization is not a result of peer influence. Scratchers are not influenced by their peers’ consumption patterns – if peers tend to ‘favorite’ projects from a specific community of projects, it does not influence Scratchers to develop a similar preference in future.
- For peer influence analysis and validations conducted above, we use exact matching to obtain control or counter-factual group where Scratchers are similar to those in the treated group except for the peers’ behaviour under study; this helps to minimize bias to a large extent, compared to using propensity score matching [24]. By ensuring balance of peers’ characteristics (in addition to individual characteristics) we control not only for homophily [24], but also for other confounding effects including selection and endogenous processes involved in network formation [16], and own behavioural tendencies.

In the next section (Section 2.2), we describe the Scratch platform and the data we analyze in more details. We also describe the measures used for peer influence analysis here. This is followed by, in Section 2.3, a

---

<sup>3</sup>The intuition is similar to clustering products in a supermarket: products that are often consumed together are classified into the same group.

**Figure 2:** PRODUCTION AND CONSUMPTION PERSPECTIVES OF A PROJECT IN SCRATCH 2.0



(a) A project is produced by composing various sprites that have codes, images, and sound associated to them. (b) A shared project is available to other users for consumption activities like viewing, downloading, loving, and commenting.

detailed description of the methods we use to identify peer influence and the results of peer influence analysis. In Section 2.4, we investigate the causal mediating mechanism of peer influence, i.e., how Scratchers’ react in response to peers’ production popularity which enhances their own popularity in future. Finally, in Section 2.5, we make a discussion based on our findings, and provide suggestions for further research.

## 2.2 Data: Scratch Community

Data from March 2007 to March 2012 was provided by the MIT Media Lab under the Scratch Research Data Sharing Agreement [7]. It consists of various metadata, corresponding to the descriptions below, of all users and their friendship formations, and of all the projects created during this period. Hence this forms a complete data set of time-stamped users' friendship network and production and consumption of projects for the first 5 years.

We analyze users' behaviour in the Scratch community.<sup>4</sup> Users come from various countries. The platform, designed for children in schools, serves as an educational media to collectively learn programming by creating and sharing interactive objects. An interactive object created on the platform is called a *project*, which is usually an animation, game, or simulation created using the Scratch programming language (SPL) [25]. Projects are composed from animated objects called *sprites*. SPL employs drag-and-drop programming method to create projects, using Scratch Authoring Environment (SAE), by assembling basic visual elements called *blocks*. The online platform was created in March 2007. SPL has had two major development versions - Scratch 1.x (1.0 to 1.4) and Scratch 2.0 (released in May 2013). To build or edit a project in 1.x versions, users had to download the Scratch editor software (offline version) to access the SAE. Users could then (optionally) share the projects in the online community. In version 2.0, which replaced 1.x, users can access SAE both online and offline.

In the Scratch community users can (i) produce projects, (ii) consume projects, (iii) follow other users as friends, and (iv) create and comment on galleries. Such collective action in Scratch community is analogous to activities in the social media platform Facebook where contents (posts or status updates) are produced and consumed by the platform users, and users can also follow each other. Projects created (Fig. 2(a)) on the Scratch platform can be of two types - *new*, and *remix*. A new project, as is suggestive, is a fresh project created by a user and shared on the website. A remix project shared by a user is a project that is created by modifying an already existing project (new/remix) on the platform. After a project is shared by a user, it can be consumed (Fig. 2(b)) by other users on the platform. Consumption of a project on the Scratch website refers to the following interactions with the project by logged-in users: *viewing*, *downloading*, *loving*, *commenting*, and *favoriting*. Each form of consumption of a project by a user is recorded only once - the first time the user interacts with it. Views, downloads, and loves of a project are anonymous

---

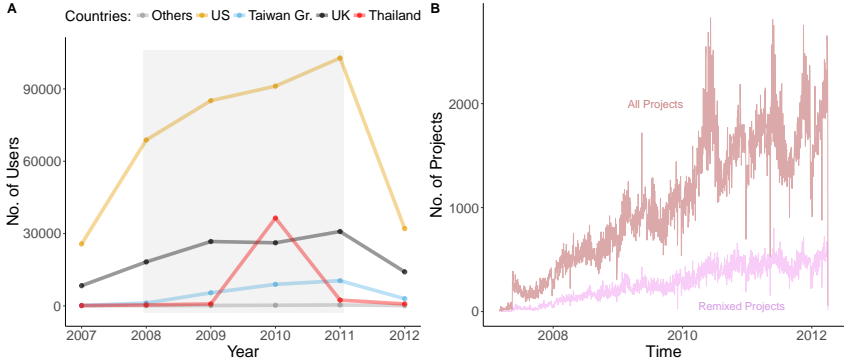
<sup>4</sup>Scratch (<https://scratch.mit.edu>) is an online educational platform created and maintained by the Lifelong Kindergarten Group at MIT Media Lab.

records, i.e, the names of the users who interacted with the project by such forms are not recorded. Friendships represent unidirectional relationships between users. A user can choose to follow any other user on the platform. Once logged in, a user can see the latest projects of the users he is following in a dedicated section. Users can also create *galleries* which are collections of projects. Users can view and comment on galleries. Projects and galleries can also be *tagged* by their creators. Tag names are not pre-defined on the platform, and new tag names are created when users tag projects and galleries with non-existing tag names. Projects and galleries can have common tags names.

Additionally, selected projects are displayed in the *front page* of the Scratch website due to various criteria (most remixed, most viewed, etc.). This selection is automated. Within each category, the three most recently added projects are displayed at any given point. There is a section on the front page for *featured projects* (three projects at a given time); projects in this section are manually added by users who are Scratch website administrators based on popularity and appeal of projects. For galleries, there are two sections on the front page, one is a section called *featured galleries*, and the other is called *studio design*. Addition of galleries to these sections are controlled by administrators. A user can at some point be assigned as a *curator* by administrator. The curator selects projects for the Scratch website's front page section labelled 'Curated By'. This section displays three recent projects selected by the curator. There is only one active curator at a time.

The dedicated section where Scratchers can see the activities of their peers in real-time is called 'What's Happening?' [26]. Here Scratchers can see the following recent activities of their peers – sharing (creation) of projects, remixing, love-its, favorites, following (users, studios). This is an important channel of information about peers' activities; if a Scratcher is following many others, he would most likely be influenced by activities of those which appear frequently via this feed. It is important to note that a Scratcher can know which projects his peers are favoriting via activity feed, however the projects which receive the favorite clicks do not show such counts on the project page. We see in Fig. 2(b) that favorites

**Figure 3: JOINING OF USERS AND CREATION OF PROJECTS**



Panel A shows the joining of 1,056,950 users during each year from March 2007 to March 2012. The evolution is grouped into major clusters of countries. Panel B shows the number of projects created daily upto March 2012. A total of 1,928,699 projects were created during the period.

(star symbol) count are visible, however this is for the latest version of Scratch. During the period 2007-12 for which data is available, SPL versions 1.x were in place, and favorites count was not visible on the project page. The love-it counts (and all other forms of consumption except favorites) on the other hand are shown on the project pages, and is public information; this forms the difference between favorites and love-its.

A schematic representation of interactions on the platform as discussed above is shown in Fig. 23. Technical details about data quality (missing data, possibly spurious data) are documented in [7]. Wherever required for this study, we discuss the data quality during our analysis.

### 2.2.1 Users, Projects

1,056,950 users joined the Scratch community in the first 5 years (Fig. 3A) and 1,928,699 projects were created during this period (Fig. 3B). The clusters in Fig. 3A are obtained using K-means clustering with five clusters; Taiwan Group is a set of nine countries. The most distinguishing

trends are born by US and UK, and there was a spike in the number of users from Thailand during 2010. We mention some statistics to describe active users on the platform. (i) There are 427,110 users with at least one non-anonymous activity. Since anonymous records include only certain forms of consumption of projects (views, downloads, loves), a large fraction of users in the data are pure consumers. (ii) There are 195,649 users who have interacted (at least one kind of recorded activity) in more than one month (months need not be consecutive). This value does not include users who might have interacted more than one month, but their interactions each month is not recorded (i.e., they only viewed, downloaded, or loved projects). (iii) There are 304,793 users (28%) who created at least one project. This is the sub-population that contributed to the 2 million projects during the five years.

### 2.2.2 Measures used for analysis

We mention the measures used for production popularity and consumption preference.

**Production Popularity** There are various observable measures that convey information about popularity of a project. These include counts of love-its, downloads, and comments. (These measures for a project are highly likely to be determined at a narrow time interval, most likely after a user has viewed and interacted with the project.) Favorites count is not observable as a consumption statistic on project page, multiple comments can be made on a project by a single consumer, and downloads count has data issues (the count is supposed to be one per user, but multiple count was found for some users). So we choose love-it as our measure of popularity; one consumer can love a project once only. Although there is no platform-specific measure for a project's quality, the love-its received on a project supposedly captures the quality of the project, as assessed by consumers who viewed the project.

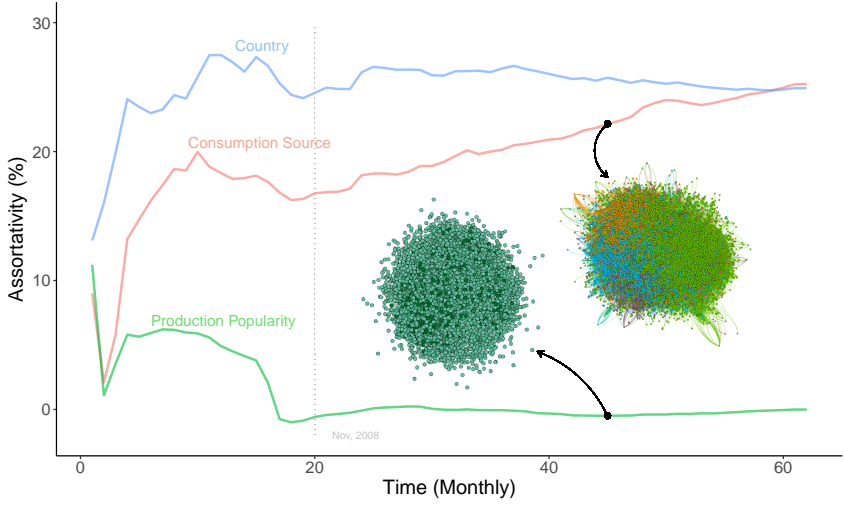
**Consumption Specificity** For consumption preference of a Scratcher, we look at the consumption source (a group of projects) from which he consumes (favorites) the most. We identify several consumption sources using unsupervised learning from the network  $\mathcal{P}_{favorites}$ , where the nodes represent projects and an edge between two nodes represents the number of users who have favorited both the nodes. The choice of algorithm to detect consumption sources does not affect the nature of our results, as presented in Section 2.3. We use the five big communities detected in  $\mathcal{P}_{favorites}$  as the set of all sources. A.2 contains details about consumption sources and about consumption behaviour, in general, on Scratch platform.

### 2.2.3 Assortativity in Behaviour

A friendship network is formed by Scratchers following each other on the platform. This can lead to observations of behavioural similarities among Scratchers and their neighbours. We look at how two attributes – production popularity and consumption preference – are clustered in the network.

We measure clustering in behaviour using the assortative mixing coefficients [27], considering numeric and categorical values for production and consumption behaviours respectively. The evolution of the assortativity values are shown in Fig. 4. The coefficients are significant at 1% level, tested using null model obtained by random shuffling of all edges in the friendship network. (As a comparative reference, the evolution of clustering based on similarity of Scratchers’ countries is also shown.) There is no trend for clustering according to production popularity and there is high clustering based on consumption preference. Fig. 4 also shows subsets of the network in December 2010. Edges of the subgraph for production popularity are not shown (for clarity of visualization); snapshots are plotted using ForceAtlas2 algorithm [28], so nodes in closer vicinity represent closer neighbours. Scratchers having same consumption sources are clustered and the pattern is dense. Scratchers with similar production popularity are however not clustered – since the

**Figure 4: ASSORTATIVE MIXING IN FRIENDSHIP NETWORK**



The evolution of assortativity coefficients for production popularity and consumption source. For a given month, assortativity is calculated using all edges in the network upto that month. Assortativity for country attribute is shown as a reference. The evolution pattern seems to be stable after November 2008. Insets show subsets of the network at December 2010 to visualize the assortativity values.

measure has many values, we rescaled the colors to have more weight on high values – users with high values of popularity are not clustered and are distributed throughout.

Observations of behavioural clustering in network (Fig. 4) can arise due to several mechanisms, including peer influence. To identify if peer influence actually exists, other mechanisms that induce clustering in behaviour need to be controlled [29, 30]. We investigate this in the following section.



## 2.3 Peer Influence Analysis

Here we study the effect of Scratchers' friendship network on their production and consumption of projects.

For production behaviour, we study the popularity effect – whether the popularity of projects created by a Scratcher increases (or decreases) if his peers' projects are popular. For consumption behaviour, we study the preference effect – if peers of a Scratcher consume (by favoriting) projects from a certain source, does the Scratcher tend to consume projects from the same source in future?

We first describe our methodology, and next we present the results.

### 2.3.1 Methods

**Potential Issues** To set the terminology, the focal node or focal actor in a network is called the ego and ego's immediate neighbours are called alters or peers. We want to infer if peers' behaviour influences ego's behaviour. Individuals in a dynamic social network may interact due to many reasons, leading to a simultaneous evolution of network and behaviour of individuals. Some of these reasons established in the literature include [31, 32, 33, 34, 16] exogenous network formation due to homophily and selection, endogenous network formation due to reciprocity and transitivity, peer influence on behaviour, own influence on behaviour, and contexts that lead to certain network-behaviour dynamics. Therefore estimating peer influence using a cross-sectional observation is prone to effects coming from other unwanted reasons, some of which can be of confounding nature. In estimating peers' influence on behaviour, confounding factors are those factors that affect both the ego's behaviour and the peers' behaviour under study. All reasons mentioned above are potentially confounding. An unbiased estimation of peer influence therefore requires control over such confounding factors. Dynamic observations facilitate the separation of co-evolution issues - for example, homophily and influence.

**Scenario & Assumptions** A Scratcher’s primary activities are producing and consuming projects and galleries. The Scratcher (ego) has the option to follow the creators (peers) of projects – which he likes during random browsing, which he likes in galleries, or interesting projects that appear on the front page of the website. Through activity feeds, the ego knows about consumption (love-its, favorites) and production (projects sharing, remixing) activities of his peers. So we can expect that the ego’s activities might be influenced by his peers, in addition to his own tendencies to produce and consume projects. At a given time, the ego can also browse through the projects of his peers to see the production and consumption statistics. These statistics, like comments count on a project, are the aggregate comments the project has received till this moment. So at a given time, we can assume that the Scratcher has knowledge about his peers and aggregate statistics of their activities (total projects, total loves received, etc.).

In measuring peer influence at a given time, we assume that the Scratcher is influenced only by the aggregate activities of his peers upto this time and not by the history of such activities. It is very unlikely that a Scratcher remembers the exact history of his peers’ activities. For example, consider a Scratcher with just one peer, and we are interested to investigate the peer influence of projects count: we assume that the total projects produced by the peer upto a given time influences the Scratcher on how many projects he produces next, and he is not particularly influenced by the exact number of projects his peer produced during the last week (or any particular historical period in general). This behaviour is termed mathematically as Markovian. Markov nature of decision making is a very plausible assumption in the scenario of Scratch community. This property has been widely adopted in the social networks literature - for example, in stochastic actor oriented models [14], future decision of network or behaviour change made by an actor is conditioned on the network and behaviour in the present state. Essentially, under Markov assumption all variables of interest are represented as state variables at the time peer influence is evaluated.

**Quasi-experiment** We define peer influence of a behaviour  $b_{peers}$  at a time  $t$  on ego's behaviour  $b_{ego}$  at time  $t + j$  ( $j = 1, 2, \dots$ ) as the exogenous influence of peers' state of behaviour (known to ego) at  $t$ ,  $b_{peers}^t$ , on ego's state of behaviour at  $t + j$ ,  $b_{ego}^{t+j}$ . State variables at  $t$  summarize behaviours upto time  $t$  and form the basis of ego's decisions at  $t$  (Markov nature). To measure peer influence, treatment status is assigned using a binary variable  $Tr^t$  which is based on a threshold value of  $b_{peers}^t$ . All Scratchers (entire population) at  $t$  are distributed into two groups: treated and control; Scratchers in the treated group ( $Tr^t = 1$ ) have high values of  $b_{peers}^t$ , and those in the control group ( $Tr^t = 0$ ) have low values of  $b_{peers}^t$  and serve as the counterfactual. At this point, treatment is likely to be correlated with several confounding variables, and so treatment effect estimates would be biased. So we obtain a subset of the population at  $t$ , by matching exactly on confounding variables, such that treatment can be justified to be randomly assigned across treated and control groups in the subset. In this sub-sample, having controlled for possible confounding effects, we capture the effect of treatment on change in behaviour of treated group ( $\Delta b_{ego}^{t \rightarrow t+j} = b_{ego}^{t+j} - b_{ego}^t \mid Tr^t = 1$ ) and compare it with the counterfactual effect ( $\Delta b_{ego}^{t \rightarrow t+j} \mid Tr^t = 0$ ). Peer influence at  $t$  is thus measured as the difference of the future changes in behaviour  $b_{ego}$  across treated and control groups. This forms the basis for peer influence estimation under a quasi-experimental setting. Below we present the empirical implementation and discuss the validity of our method.

**Implementation** We employ an ego-centric regression framework to assess, at time  $t$ , the effect of being treated on ego's future change in behaviour.

$$\begin{aligned}
\Delta \mathbf{b}_i^{t \rightarrow t+j} &= \alpha^j + \beta_{peer}^j \mathbf{Tr}_i^t \\
&+ \underbrace{\beta_1^j \mathbf{N}_i^t + \beta_2^j \mathbf{X}_i^t}_{\text{confounders are balanced across } Tr^t \in \{0,1\}} \\
&+ \epsilon_i^{t \rightarrow t+j}, \quad j = \{1, 2, 3, \dots\}
\end{aligned} \tag{2.1}$$

$\Delta b_i^{t \rightarrow t+j}$  is the change in behaviour  $b$  of ego  $i$  from time  $t$  to  $t + j$ . All explanatory variables represent behavioural state at  $t$ , measured as an aggregate operation on observed behaviour upto  $t$ . For example, to represent the comments behaviour of ego at  $t$ , we use the total comments made by the ego upto  $t$  as the state variable at  $t$ .  $Tr_i^t$  is the treatment variable – the variable of interest that represents peers’ behaviour at  $t$ . It is a binary variable – values 1 and 0 are assigned to egos in the treated and control groups respectively. Under absence of selection bias,  $\beta_{peer}^j$  represents the average treatment effect on change in future behaviour. Treatment status  $Tr_i^t$  for ego  $i$  can change over time (i.e., a Scratcher who is in the treated group today can be in the control group at another time) because the ego is assigned to either treated or control group based on peers’ behavioural state at  $t$ . Hence estimates  $\beta_{peer}^j$  are conditional on time  $t$ .  $N_i^t$  represents ego’s network variables at time  $t$ . In general, it can incorporate information of the entire network of ego upto neighbours at any distance. However, for practical purpose it is sufficient to include characteristics of ego’s local network (immediate neighbours) only – structural properties of ego’s network (e.g., out-degree, in-degree, reciprocity), various behaviours of peers (excluding the behaviour represented by the treatment variable), and structural properties of peers’ local network.  $X_i^t$  represents various characteristics of the ego at time  $t$ . It includes the dependent behaviour  $b$  under study as well to capture auto-correlation of behaviour, or in other words, ego’s own tendency. In this study we use monthly windows –  $t$  is the time at a month’s end,  $t + 1$  is the time at the end of next month,  $t + 2$  is the time at the end of 2 subsequent months from  $t$ , and so on. We have included all potential confounders in the sets of covariates  $X^t$  and  $N^t$  which could be observed in the available data. Potential confounders consist of those variables which, conceptually,<sup>5</sup> can affect both  $Tr_i^t$  and  $\Delta b_i^{t \rightarrow t+j}$ , and hence neglecting such variables can bias the peer influence estimates.

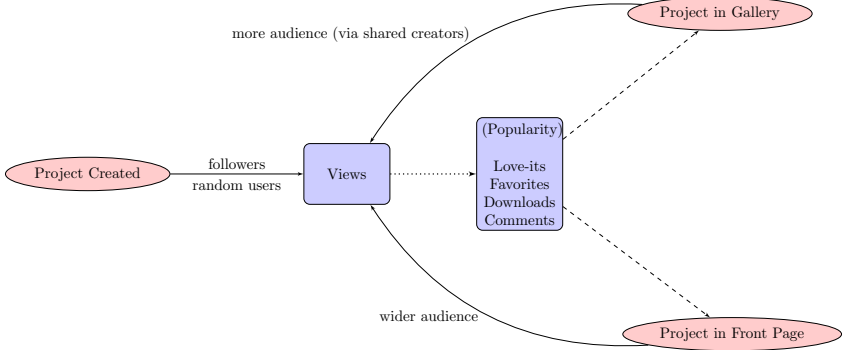
---

<sup>5</sup>For example, ego’s own behaviour at time  $t$  is conceptually a determining factor of his peers’ behaviour at  $t$  (possibly due to homophily), and also his future behaviour. Hence neglecting such a variable in  $X^t$  can bias the peer influence estimates upwards. However it is noteworthy that, in the data, all conceptual confounders need not be statistically significant.

**Preprocessing for covariates balance** Exact matching is used as a preprocessing step prior to estimating (2.1) in order to achieve balance in confounding variables across treated and control groups. Regression analysis following the matching step leads to statistically consistent estimates [35]. We implement one-to-many exact matching, in which each treated unit is matched to multiple units in the control group having exactly the same values of the matched variables [36, 35]. Each matched control unit has weight proportional to the number of treatment units to which it is matched, and the sum of the control weights is equal to the number of uniquely matched control units. Unmatched units have weights equal to 0, and matched treated units have weight 1. The regression analysis that follows matching uses weights corresponding to each unit produced during matching stage [35]. Exact matching costs data, so we exploit high correlations among variables to obtain balanced samples by matching only on subsets of (important) confounders. Eventually, for regression analysis, we use samples which produces the best balance (reduces bias in regression estimates), and also retains a good sample size (reduces variance of regression estimates). Variables which remain unbalanced are included at the regression stage as controls. Balance in model variables in (2.1) is assessed by difference in weighted means of variables across treated and control groups. Balance is assessed on all model variables, including those that are excluded from matching analysis. For a given matched sample, with a good balance of the post-matched variables across treated and control groups, regression estimates are not supposed to change dramatically across different models.

**Summary of steps involved** We summarize the practical steps involved in estimating peer influence using the method presented above. (i) determine model variables, i.e., all potential confounders (ii) dichotomize treatment, if needed (iii) determine selection into treatment, i.e., statistically significant confounders (iv) match exactly on (subset of) confounders to achieve balance (v) estimate model using OLS.

**Figure 5: ACCUMULATION OF LOVE-ITS**



The pathways leading to accumulation of love-its. After creation and sharing of a project, love-its on the project can accumulate from views by (1) followers of its creator, (2) users who view it when it appears on the front page (after it becomes popular due to various factors), (3) followers of shared creators of a gallery where the project appears, and (4) random views on the project due to users' browsing.

**Internal Validity** We want to obtain unbiased estimates  $\beta_{peer}^j$  of the treatment effect in (2.1). Since we do not expect reverse-causality issues, selection bias is the most important source of bias. This arises due to confounding factors that affect both the treatment and future change in ego's behaviour, and are not affected by the treatment itself or by anticipation of treatment. Same intensity of selection bias across treated and control groups can justify a random assignment of treatment, and minimize alternative mechanisms. (a) Exact matching as a preprocessing step and including controls in the regression stage help to minimize selection bias due to observable factors. We control for exogenous network formation processes, homophily and selection, by accounting for balance in ego's characteristics and peers' characteristics [16]. Endogenous network formation processes like general tendency to follow Scratchers (out-degree) and tendency to follow one's followers (reciprocity) can be confounding, so we control for such factors as well. Change in future behaviour can also depend on the level of behaviour at  $t$ , and this is a major confounder

because Scratchers in the treated group are more likely to have higher behavioural levels because of correlation between peers' behaviour and ego's behaviour. We therefore always include this factor in all matching analysis; the sub-samples have exact levels of behaviour across treated and control groups before the onset of change in behaviour. (b) Comparison of future change in ego's behaviour with a counterfactual group takes care of selection into treatment due to unobservable factors, as long as such factors are time-invariant. (c) The treatment variable is dichotomized [35] before matching analysis according to certain thresholds that suggest high or low levels of behaviour. We conduct analysis for various thresholds to ensure that peer influence estimates are not extremely sensitive to such choices of thresholds.

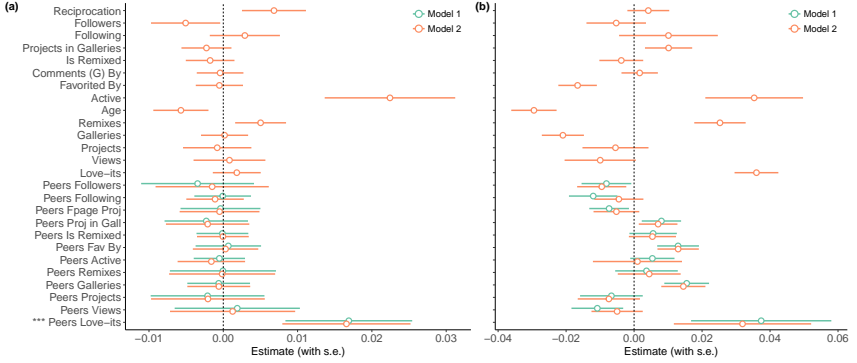
### 2.3.2 Results

For estimating peer influence on production popularity and consumption specificity, we follow the steps mentioned in previous section (Methods). We found presence of statistically significant peer influence only for production behaviour. In the remaining part of this section, we provide empirical details of peer influence for production behaviour and an outline of the empirical study for consumption behaviour.

#### Production Popularity

A Scratcher accumulates love-its on a project he created when another Scratcher (consumer), either his follower or a random user, who views the project finds it interesting (most likely due to project quality) and clicks the love-it button. If the project receives a lot of attention (as inferred by love-its, comments, downloads, etc.), it can be selected to appear on the front page. This selection can be system-based or by admins. The project can also appear in some galleries. These would most likely increase viewership for the project, and the project is subject to more love-its. Fig. 5 shows a schematic diagram of the process of accumulation of love-its. All factors that affect the quantity and quality of projects created by a Scratcher, and the total views on all his projects are predic-

**Figure 6: PEER INFLUENCE ONE MONTH AHEAD**



Estimates of regression (2.1) for  $t = \text{Dec 2010}$  and  $j = 1$  using samples obtained by matching on (a)  $X$  variables and (b) both  $X$  and  $N$  variables. In (a) and (b), Model 1 includes only peers' attributes ( $N$ ), and not egos's attributes ( $X$ ), as controls. The effect of peers' production popularity (Peers Love-its) on Scratchers' change in production popularity next month,  $\beta_{peer}^1$ , is significant at 1% level. See Table 8 for variables description, Table 11 for details of matched samples, and Table 12 for details of estimates.

tive of his popularity (measured as total love-its). In reference to model (2.1), these factors are of types  $X_i^t$  (ego's attributes) and  $N_i^t$  (ego's local network structure, peers' observable characteristics, peers' local network structure), and are selected according to the criteria mentioned in previous section (Methods). The measures of covariates  $X_i^t$  and  $N_i^t$  represent the respective behavioural states at the time  $t$  of peer influence evaluation (see Table 8).

The treatment variable of interest is popularity of peers' projects at  $t$  ( $Tr_i^t$ ); since it is not a binary measure, we dichotomize as

$$Tr_i^t = \begin{cases} 1 & PQ_i^t \in (c_{min}, c_{max}] \\ 0 & PQ_i^t \in [0, c_{min}), \end{cases} \quad (2.2)$$

where  $PQ_i^t$ , the sum of all love-its on all projects upto  $t$  of all peers of  $i$ , is the measure for peers' popularity for ego  $i$  at time  $t$ , and  $c_{min}$  and  $c_{max}$  are self-chosen values representing minimum and maximum threshold values respectively. (We shall see later how the chosen thresholds affect

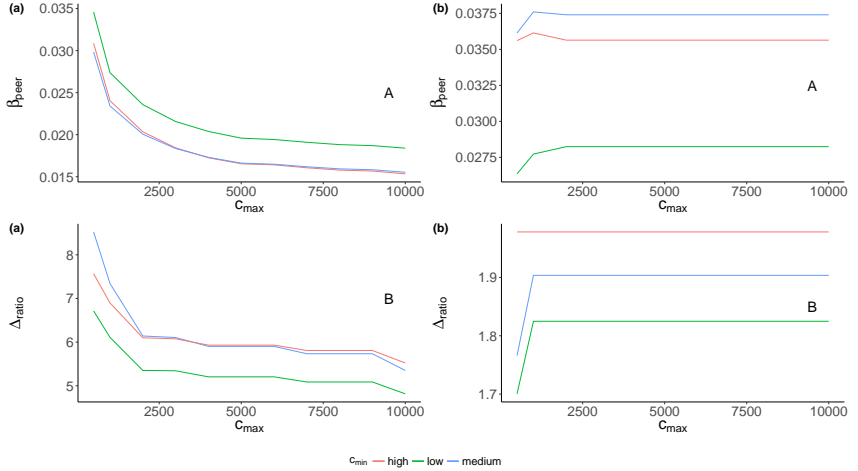


the results.) Our dependent variables of interest are future changes in production popularity of ego  $i$ :

$$\Delta b_i^{t \rightarrow t+j} = b_i^{t+j} - b_i^t, j = 1, 2, \dots,$$

where  $b_i^s$  represents the total love-its accumulated by all projects of ego  $i$  upto time  $s$ . Peer influence estimation is conditional on time  $t$ ; to have sufficient observations, we begin by analyzing in the stable period of the data: the month of December, 2010. ( $t$  represents the end of Dec, 2010 and  $t + 1$  is the end of Jan, 2011.) We use the median value of peers' popularity at  $t$  as  $c_{min}$  and the maximum value of  $PQ^t$  as  $c_{max}$ .

**Figure 7: ROBUSTNESS CHECK FOR PEER INFLUENCE ONE MONTH AHEAD**



Effect of varying threshold values of  $c_{min}$  and  $c_{max}$  on estimates evaluated on samples obtained by matching on (a)  $X$  variables, and (b) both  $X$  and  $N$  variables. ( $t = \text{Dec } 2010$ ,  $j = 1$ ) Panels A: Estimate of  $\beta_{peer}$ , controlling for both  $X$  and  $N$  variables in the regression. Panels B: Ratio of average of future changes in production popularity of treated and control groups.

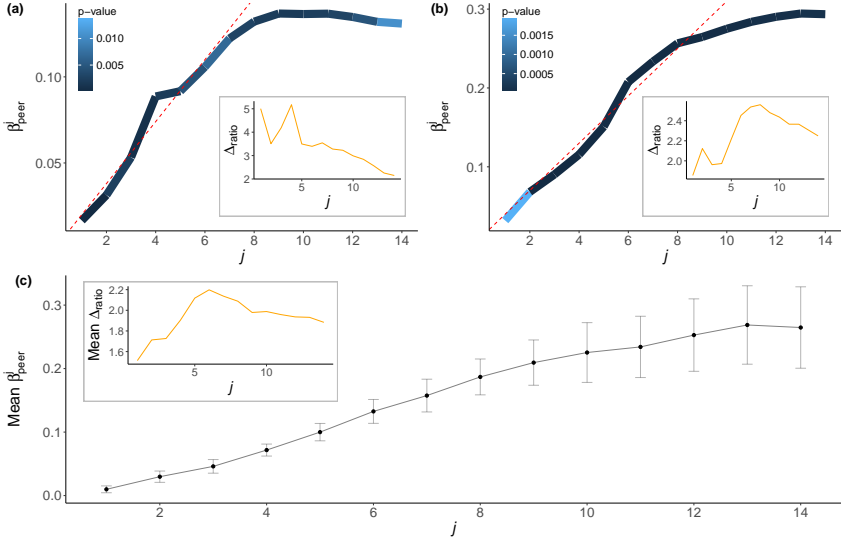
In Table 10 (columns 1,2), we see that almost all variables (included in model as potential confounders) are valid model variables. To learn which of the potential confounders are important, i.e., affect treatment

variable and hence can lead to selection bias, we perform logistic regressions of treatment  $Tr_i^t$  on covariates  $X_i^t$  and  $N_i^t$ . As shown in columns 3 and 4 of Table 10, we perform two such regressions: first with only  $X_i^t$  variables and next with all variables. Having determined the significant confounders from this analysis, we perform exact matching on (all or subset of) such confounders across treated ( $Tr_i^t = 1$ ) and control ( $Tr_i^t = 0$ ) groups with an immediate goal to have balance of all variables – includes all potential confounders, irrespective of their statistical significance in logistic regression – across both groups. Balance is determined by the difference in weighted average values of the variables in each group. We obtain two reduced datasets:

1. sample obtained by exact matching only on all  $X_i^t$  variables that are significant (Table 10, col. 3). This is done to (i) compare our method with Aral [24], which matches on all individual (ego) characteristics, and (ii) show that our results are robust to matching strategies that produce good balance,
2. sample obtained by exact matching on a subset of all ( $X_i^t$  and  $N_i^t$ ) variables that are significant (Table 10, col. 4). Matching exactly, especially with  $N_i^t$  variables was found to be very costly, and so only few variables were used. Among all combinations of variables we investigated, we present the one with best balance. This sample is less biased than the one in (a), and also has more observations because matching is performed on less covariates.

We show balance for all variables (difference in averages of variables by treatment groups) in Table 11. Although we do not use, we show balance produced by matching via propensity score method, as used by [24]; this method produced almost no improvement (as compared with the original imbalance in the full sample) in covariates balance. Having balanced samples, we can assume we are in a scenario where treatment (high popularity of peers) has been randomly assigned to each Scratcher (ego). The quantitatively small imbalances that still remain for some covariates are controlled during the regression stage.

**Figure 8: PERSISTENCE OF PEER INFLUENCE**



(a) Estimates of  $\beta_{peer}^j$ , the effect of peers' production popularity at  $t = \text{Dec } 2010$  on production popularity at future periods  $t + j$ , for varying  $j$ . Inset shows the ratio of future changes in production popularity of treated and control groups. Sample in Table 11(a) obtained by matching egos' characteristics  $X^t$  is used for evaluation. (b) Same estimators as in (a), using sample in Table 11(b) obtained by matching individual and peers' characteristics. (c) A general persistence curve, showing the average  $\beta_{peer}^j$  during July-Dec 2010, i.e., for each  $j$ -periods ahead influence, the plot shows its average value calculated for each month during July-Dec 2010. Error bars are scaled standard deviations of  $\beta_{peer}^j$ . Inset shows the ratio of mean future changes in production popularity of treated and control groups. Sample in Table 11(b) is used.

Fig. 6 shows the estimated coefficients for (2.1) with  $j = 1$ , i.e., the effect of peers' production popularity on Scratchers' production popularity the next period. Estimates are shown separately for two cases, corresponding to the two reduced datasets. For each matched sample, we see that the estimated model coefficients are stable across various specifications and the peer influence coefficient  $\beta_{peer}^{j=1}$  is positive. In both cases,  $\beta_{peer}^1$  is significant at 1% level. Details of regressions corresponding to (a) and (b) in Fig. 6 are available in Table 12. The identification of  $\beta_{peer}^1$  re-

lies on random assignment of observations to treated and control groups. Achieving a good balance of covariates across both groups and including covariates as controls in regressions minimizes selection bias to a large extent. However, another source of non-random treatment assignment lies in the definition of  $Tr_i^t$  variable which depends on the chosen values of  $c_{min}$  and  $c_{max}$ . The results in Fig. 6 use one pair of values; so we need to check whether peer influence estimate remains significant and how it varies when threshold values  $c_{min}$  and  $c_{max}$  change. Fig. 7 shows this robustness analysis. Since  $\beta_{peer}^1$  is the primary coefficient of our interest (peer influence), we plot these estimates for changing threshold values, as shown in panels labelled A. All estimates of  $\beta_{peer}^1$  are positive and significant at 1% level. For each value of  $c_{min}$ , peer influence estimate tends to decrease with increase in  $c_{max}$ . Panels labelled B show the ratio of weighted means of outcome  $\Delta b^{t \rightarrow t+1}$  variables in treated and control groups, a measure of relative comparison of outcomes without any post-matching adjustment for confounders.

Above, we provided details of the peer effect of production popularity in the immediate period ( $j = 1$ ) using the network existing at the end of December 2010 ( $t$ ). Popularity of peers' projects at time  $t$  might influence the popularity of a Scratcher's (ego) projects in subsequent periods as well; in model (2.1) this corresponds to values of  $j$  as 2, 3, and so on. To see if there is persistence of peer influence in subsequent periods, we estimate  $\beta_{peer}^j$  for various  $j$  by changing the dependent variables in (2.1). Since we are looking at the effect of peers' popularity at time  $t$  on period  $t + j$ , the treatment assignment  $Tr_i^t$ , and hence the reduced datasets obtained by exact matching, remains the same as in the analysis for Fig. 6 and Fig. 7. The results for persistence of peer effect are shown in Fig. 8(a) and Fig. 8(b). The effect of peers' popularity at  $t$  on Scratchers' (ego) production popularity at a future period  $t + j$  increases with subsequent periods. The rate of increase tends to steep up during the middle term and tends to flatten out in the long term, thereby creating a S-shape for the structure of persistence curve. This shape is more prominent in Fig. 8(b), compared to 8(a), where the balance in underlying matched sample is better (Table 11).

**Table 1: HETEROGENOUS INFLUENCE (SUSCEPTIBILITY)**

	Immediate (j=1)	Medium Term (j=6)	Long Term (j=12)
<i>(a) Matched on X</i>			
<i>2nd order effect</i>			
$\sigma$ (Peers' Popularity)	(0.019 *** , 0 , 0 )	(0.085 ** , -0.02 , 0.04 )	(0.111 ** , -0.07 , 0.1 )
<i>Activity Frequency</i>			
Active	(0.006 , 0.01 ** , 0.03 ***)	(0.034 , 0.16 ** , 0.24 ***)	(0.05 , 0.27 ** , 0.29 ***)
Age	(0.018 *** , 0 ** , 0 )	(0.128 *** , -0.02 ** , -0.01 )	(0.151 *** , -0.03 *** , -0.01 )
<i>Producer Type</i>			
Developer	(0.014 *** , 0 , 0.16 ***)	(0.104 *** , 0.28 , 0.07 )	(0.13 *** , 0.38 * , 0.25 )
Free-style	(0.017 *** , 0.08 *** , -0.01 )	(0.056 , 1.44 *** , 3.51 ***)	(0.082 * , 2.65 *** , 3.59 ***)
<i>(b) Matched on X, N</i>			
<i>2nd order effect</i>			
$\sigma$ (Peers' Popularity)	(0.036 *** , 0 * , 0 )	(0.207 *** , 0.01 , 0 )	(0.303 *** , 0.01 , -0.03 )
<i>Activity Frequency</i>			
Active	(0.016 , 0.03 *** , 0.02 )	(0.04 , 0.16 *** , 0.24 ***)	(0.038 , 0.24 *** , 0.37 ***)
Age	(0.037 *** , 0 *** , 0 )	(0.207 *** , -0.01 *** , 0 )	(0.311 *** , -0.02 *** , 0 )
<i>Producer Type</i>			
Developer	(0.021 ** , -0.05 *** , 0.23 ***)	(0.196 *** , -0.15 *** , 0.23 )	(0.28 *** , -0.25 *** , 0.2 )
Free-style	(0.032 *** , 0.05 *** , 0.01 )	(0.131 *** , 0.12 *** , 0.92 ***)	(0.198 *** , 0.26 *** , 1.14 ***)

(i) In a given row, each tuple (*a*, *b*, *c*) represents, in order, the coefficients of treatment (peers' production popularity), attribute (corresponding to the row name), and the interaction of treatment and attribute ( $\sigma$  represents variance). The coefficient of the interaction variable captures the heterogeneity of treatment with respect to the attribute.

(ii) p-values: significant at 10% level (\*), 5% level (\*\*), 1% level (\*\*\*)

Persistence curves at two different times are ideally not comparable because treatment assignment on Scratchers (ego) can vary from one period to another. So performing peer influence analysis at a time  $\tilde{t}$  different from the one we have used in above analysis ( $t = \text{Dec, 2010}$ ) requires obtaining a balanced, preferably the least biased, sample at  $\tilde{t}$  (by matching exactly on confounders that are significant at  $\tilde{t}$ , which may differ from those at  $t$ ). However, assuming that the users' behaviour to be stable over the last six months of 2010, we can assume that the selection into treatment in each of these months ( $t$ ) follows a similar pattern as we saw above. So we create reduced samples by exactly matching on the same set of variables as in column (b) of Table 11. We expect that the imbalances in other months would be more than that in Table 11 (which corresponds to month of December); however controls in the regressions help to reduce bias. Now we look at the persistence curves  $\beta_{peer}^j (j = 1, \dots, 14)$  for each of the last six months ( $t = \text{July 2010, } \dots, \text{Dec 2010}$ ); the average of these curves is shown in Fig. 8(c). For each  $j$ , we plot the average of  $\beta_{peer}^j$  estimates obtained in six different models (2.1) corresponding to six different values of  $t$ .

We saw earlier that the assortativity coefficient for projects popularity is near zero, which means that there is no observed clustering based on popularity of the projects, and this is consistent over time. However, we do find positive short term and long term peer influence, which means that a positive measure of clustering can be expected at a future time when the effect of influence has taken place. Of several probable mechanisms that might explain this, we provide a qualitative example to illustrate the main idea. Consider a situation, at time  $t$ , where an ego has four outgoing friendships, two of which have higher production popularity than him and the other two have lower popularity than him. This suggests that ego's local network is not assortative based on this behaviour. During time  $t$  to  $t + 1$ , the ego receives a small increase in its projects' popularity due to influence of his peers. However since his friends also undergo similar change (by influence from their peers), the relative values of popularity in ego's local network at  $t + 1$  remains the same as in time  $t$  – two friends have better popularity and other two have lower

popularity, suggesting a zero value for assortative mixing once again.

Next we study if particular Scratchers, due to their own nature or local network characteristics, are more susceptible to influence from their peers compared to other identical Scratchers. For this, we use an interaction term (of the desired attribute) with the treatment variable in regression model (2.1), controlling for all confounders as before. The results are shown in Table 1; in each tuple, the first, second, and last elements correspond respectively to the estimated coefficients of treatment variable, desired attribute, and the interaction of treatment and attribute. The first attribute is a network characteristic: variance of peers' production popularity; we see that the interaction term is not significant, i.e., an ego in the treated group with two peers having average production popularities will be influenced in the same way if one of his peers had a high popularity and the other had a low popularity. So a treated Scratcher is not influenced by specific peers (in general), rather the influence stems from the overall production popularity of his local environment. Not shown here, we tested for several other attributes of peers (total remixes, favorites, projects in front page, etc.) and found no evidence of influence heterogeneity. So, peers' behaviour does not seem to create extra susceptibility, which seems intuitive, because incorporating more influence in comparison to identical others should arise out of individual traits. We see that active Scratchers are influenced more than if not active, and the effect on future periods is increasing. (Active users are those who have interacted on the platform for at least one month. Non-active users can be of two types – those who joined much before but interacted less than one month, and those joined just one month prior to  $t$ , i.e., during December 2010.) Further activity frequency does not have a significant effect among treated Scratchers, as seen from the coefficient of age. So it seems that, on average, a minimal duration of interaction (one month), either initially or somewhere during the lifetime, on the platform makes Scratchers more susceptible to the production popularity of their peers. Remixing is an important characteristic in the learning process in Scratch community. Based on an individual Scratcher's preference to create remixed projects more frequently than creating original projects,

we have segmented Scratchers into 3 types: innovators, free-style users, and developers, in increasing order of preference to produce remixed projects. The group of users who tend to produce remixed projects most often are termed developers. We do not consider innovators in subsequent analysis because there is a possibility for a remixed project to be presented as an original project on the Scratch platform. (See A.1 for details on segmenting users based on the type of projects they produce.) We see in Table 1 that developers are more susceptible to be influenced in the very immediate period, but not in the medium or long term, compared to other producers whose peers have high production popularity but they are either free-style producers or innovators. Most probably, it is by the nature of remixing – if a developer sees popular projects of his peers, he builds on top of it to have new projects in the next period and gain popularity, but he is not influenced by today’s production of peers to create projects in the subsequent periods. On the other hand, influence of peers’ popularity on free-style producers takes effect in the medium to long term and is very significant. So having traits of innovation, i.e., creating new projects, leads to additional influence from production popularity of peers in the long run.

## Consumption Preference

The next behaviour that we examine for peer influence relates to consumption. In Fig. 4 we saw that Scratchers having the same source of consumption tend to cluster in the friendship network. We investigate to what extent this can be explained by influence from peers, i.e., if peers tend to favorite (consume) projects from certain source, does the ego also tend to increase consumption from the same source?<sup>6</sup> This investigation is relevant because the ego knows about his peers’ favoriting patterns via activity feeds (comments made are not visible to followers via activity feeds), but does not know the ‘source’ of consumption because the sources have been identified by us by clustering projects over the observed data of the entire duration. (We suggest readers to refer to Section

---

<sup>6</sup>We use favorites to understand consumption because peers’ favorites are visible as activity feeds.



2.2.2 for the measure of consumption specificity and A.2 for further details.) Evidence of peer influence in this case would imply that tastes or preferences for consumption of ego are not static and can be affected by peers' preferences.

We consider the group of projects in  $c_2$ , the biggest community in the  $\mathcal{P}_{favorites}$  network. We examine whether ego increases his consumption of projects from  $c_2$  group if his peers mostly consume projects from  $c_2$  group. In terms of model (2.1),  $\Delta b_i^{t \rightarrow t+j}$  is the change in consumption of  $c_2$  projects from  $t$  to  $t+j$ , and  $Tr_i^t$  for ego  $i$  is 1 only if more than 50% of his peers have consumed projects from  $c_2$  group the maximum time (upto  $t$ ). Controlling for various confounding factors, we did not find significant coefficient for  $\beta_{peer}^j$ . So we can not conclude that Scratchers are influenced by their peers' consumption interests. The most likely reasons for the observed clustering in Fig. 4 therefore seems to arise out of contextual friendship formation among the users, context being the position in the network after initial interactions on the platform, which is followed by consuming projects within large communities locally. Since each consumption basket/source contains projects of similar themes (as described in A.2), it probably explains why Scratchers are not influenced by the consumption patterns (favorites) of their peers.

## 2.4 Mechanism of Peer Influence

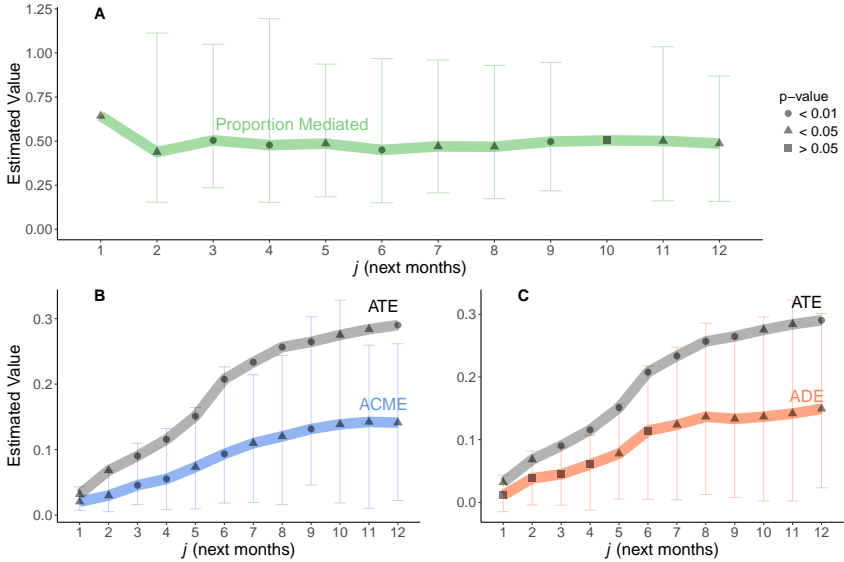
In the previous section we saw that production popularity of peers has a positive effect on the outcome of Scratchers' future production popularity (Fig. 8). However this finding does not suggest *how this effect mediates to outcome* – especially, whether any particular changes in activities made by Scratchers in subsequent periods due to popularity of peers' projects affects their future production popularity. The aggregate love-its accumulated by a Scratcher upto a certain time  $t$  depends on the projects created by him and the views received on his projects upto that time. The change in production popularity in next periods, i.e., during  $t$  and future times  $t+j$ , can therefore arise due to:

- (*Channel 1*) change in the number of projects created in next peri-

ods, which leads to new views and the possibility to have more love-its, and

- (*Channel 2*) change in the number of views in next periods on projects already created by time  $t$ , which can lead to more love-its.

**Figure 9: PEER INFLUENCE CHANNEL FOR PRODUCTION POPULARITY: CREATION OF PROJECTS**



(A) The proportion of peer influence in production popularity, as estimated in Fig. 8(b), which is mediated via Scratchers' creation of new projects in future periods  $j$ . 95% CIs are shown for each estimate, except at  $j = 1, 10$  where the upper boundary is more than 1.25. (B, C) The decomposition of the total average effect of peer influence (ATE) into the primary channel of creation of new projects (ACME), and secondary channel which includes all other pathways (ADE). 95% CIs are shown for ACME and ADE estimates.

Of these two (major) mechanisms of peer influence, the outcome (change in popularity next period) caused via creation of new projects is particularly noteworthy. This mechanism gives better insight into the Markov decision-making nature of Scratchers – upon having popular peers, they are influenced to create more projects (and possibly of better quality)

which enhances their popularity in future periods. We denote  $b_i^t$ ,  $Tr_i^t$ , and  $M_i^t$  as the production popularity, treatment assignment, and total projects respectively at time  $t$  for Scratcher  $i$ . Our goal is to disentangle the effect mediated via creation of new projects (channel 1):

$$\begin{array}{ccccc} Tr_i^t & \xrightarrow{\text{decision}} & \Delta^j M_i & \xrightarrow{\text{views}} & \Delta b_i^{t \rightarrow t+j}(Tr_i^t, \Delta^j M_i) \\ \text{peers' popularity} & & \text{new projects} & & \text{gain in popularity} \end{array}$$

from other mechanisms of treatment effect (channel 2), where

$$\Delta^j M_i \equiv \Delta^j M_i(Tr_i^t) := \Delta M_i^{t \rightarrow t+j}(Tr_i^t) = M_i^{t+j}(Tr_i^t) - M_i^t$$

is the total projects created by Scratcher  $i$  during  $t$  and  $t+j$  (the mediating variable of interest), and is a function of peers' production popularity at  $t$  ( $Tr_i^t$ ).

To do so, we employ model-based causal mediation analysis [23] where the treatment variable  $Tr_i^t$  is randomized, conditional on the confounders  $X_i^t$  and  $N_i^t$  (quasi-experimental setting), and the mediating  $\Delta M_i^{t \rightarrow t+j}$  and outcome  $\Delta b_i^{t \rightarrow t+j}$  variables are observed without interventions. Estimates of  $\beta^j$  in regression (2.1) is the total average treatment effect  $ATE(j)$ <sup>7</sup>, and is the sum of average effects of treatment at  $t$  on gain in production popularity from  $t$  to  $t+j$  via all possible mediating channels.  $ATE(j)$  is decomposed into two components:<sup>8</sup>

- $ACME(j)$ : the component of  $ATE(j)$  mediated via creation of new projects (Channel 1) is called the *average causal mediation effect* and is defined as:

$$\begin{aligned} \bar{\delta}^j(Tr^t) = \mathbb{E}_i[ & \Delta b_i^{t \rightarrow t+j}(Tr^t, \Delta^j M_i(1)) \\ & - \Delta b_i^{t \rightarrow t+j}(Tr^t, \Delta^j M_i(0)) ], \end{aligned}$$

---

<sup>7</sup> $\beta^j$  in regression (2.1) estimates the average treatment effect on the treated group ( $Tr^t = 1$ ) and is equal to the average treatment effect (ATE), for the population, when  $Tr^t$  is randomly assigned.

<sup>8</sup>Note that  $ACME(j)$  and  $ADE(j)$  are defined using potential outcomes framework and hence contain counterfactual values. For example, for a Scratcher with  $Tr_i^t = 1$ ,  $\Delta M_i(0)$  is not observed because it is the number of projects he would have created from  $t$  to  $t+j$  if he were assigned  $Tr_i^t = 0$ . Counterfactual values are estimated from the data during the estimations of  $ACME(j)$  and  $ADE(j)$ .

- $ADE(j)$ : the component of  $ATE(j)$  mediated via all other mechanisms (Channel 2) is called the *average direct effect* and is defined as:

$$\begin{aligned}\bar{\xi}^j(Tr^t) = & \mathbb{E}_i[ \Delta b_i^{t \rightarrow t+j}(1, \Delta^j M_i(Tr^t)) \\ & - \Delta b_i^{t \rightarrow t+j}(0, \Delta^j M_i(Tr^t)) ].\end{aligned}$$

To contrast our investigation with the previous results, we fix  $t$  at December, 2010 and vary  $j$  from 1 to 12. Following [23, 37], we estimate  $ACME(j)$  and  $ADE(j)$  in the reduced dataset obtained by matching all variable types (Table 11(b)). Linear models (2.3) and (2.4) are used for mediating and outcome variables respectively:

$$\Delta M_i^{t \rightarrow t+j} = \gamma_0^j + \gamma_1^j Tr_i^t + \gamma_2^j N_i^t + \gamma_3^j X_i^t + \epsilon_{M_i}^{t \rightarrow t+j} \quad (2.3)$$

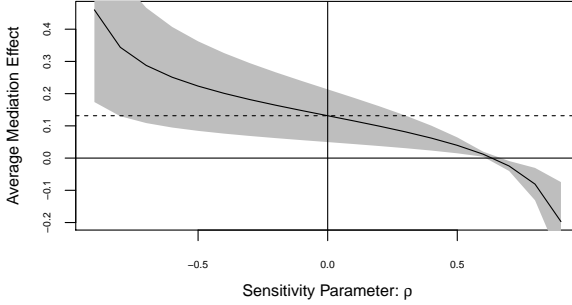
$$\Delta b_i^{t \rightarrow t+j} = \beta_0^j + \beta_1^j Tr_i^t + \beta_2^j \Delta M_i^{t \rightarrow t+j} + \beta_3^j N_i^t + \beta_4^j X_i^t + \epsilon_{b_i}^{t \rightarrow t+j} \quad (2.4)$$

$$\rho^j = Corr(\epsilon_M^{t \rightarrow t+j}, \epsilon_b^{t \rightarrow t+j}), \quad (2.5)$$

where  $\rho^j$  is the correlation between the error terms, and  $X^t$ ,  $N^t$  are the same confounding variables as used in model (2.1), for estimating peer influence on production popularity, to ensure a random assignment of treatment. The estimates and their confidence intervals are obtained using non-parametric bootstrap with percentile method.

Fig. 9 shows the estimates of  $ACME(j)$  and  $ADE(j)$  in panels B and C respectively. We see that both estimates are positive and increasing for all  $j$ . Both the effects tend to increase at a decreasing rate, and so we observe a similar additive effect for  $ATE(j)$ . Also, the S-shape for  $ATE(j)$  seems to arise from  $ADE(j)$ . The ATE curve in Fig. 9 (obtained here using non-parametric bootstrap estimation) is identical to the curve in Fig. 8(b). We performed a heterogeneity test [37] with the null hypothesis  $\bar{\delta}^j(1) - \bar{\delta}^j(0) = 0$  and concluded that  $\bar{\delta}^j(1)$  and  $\bar{\delta}^j(0)$  are statistically not different (high p-values  $\forall j$ ); so the  $ACME(j)$  curve in Fig. 9(B) holds for treated and control groups, i.e., the average mediating effect does not depend on treatment status and hence is the same for all Scratchers. For an alternate interpretation, Fig. 9(A) shows the estimated values and

**Figure 10: ROBUSTNESS CHECK FOR PEER INFLUENCE CHANNEL**



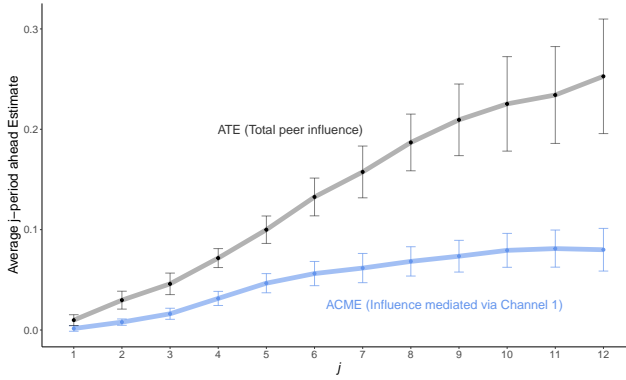
The variation of  $ACME(j = 9)$ , along with 95% confidence intervals of the estimate, with change in the sensitivity parameter  $\rho^{j=9}$  (see (2.5)). (Plots of variation for other values of  $j$  are identical to that shown here, for  $j = 9$ .) The parameter  $\rho^j$  is sensitive to the presence of confounding effects.  $\rho^j = 0$  corresponds to validity of the sequential ignorability assumption.

confidence intervals for the proportion of  $ATE(j)$  that is mediated via creation of new projects, given by  $\frac{\bar{\delta}^j(1)}{\bar{\delta}^j(1) + \bar{\xi}^j(1)}$ . Almost 40 – 50% of the effect of peers' popularity on Scratchers' future production popularity is explained by the creation of new projects in all future periods subsequent to the treatment at  $t$ .

We perform robustness check for the estimates obtained in Fig. 9. Identification of estimates  $\bar{\delta}^j$  and  $\bar{\xi}^j$  assumes sequential ignorability, a set of following assumptions: (i) exogeneity of  $Tr^t$  in models (2.1) and (2.3) conditional on  $X^t, N^t$ , and (ii) exogeneity of  $\Delta M^{t \rightarrow t+j}$  conditional on  $Tr^t$  and  $X^t, N^t$ . Assumption (ii) can be violated, even if  $Tr^t$  is randomly assigned, by post-treatment variables that affect both mediating and outcome variables in (2.4). Since this is an untestable assumption, a sensitivity analysis [23] is used to gauge the reliability of the estimates  $\bar{\delta}^j$  and  $\bar{\xi}^j$  in Fig. 9. For this  $\rho^j$  in (2.5) is used as the sensitivity parameter since variables that violate assumption (ii) will be present in both models

(2.3) and (2.4). The estimates in Fig. 9 are obtained by assuming sequential ignorability, i.e.,  $\rho^j = 0, \forall j$ . Sensitivity of  $ACME(j)$  estimate to other values of  $\rho^j$  is shown in Fig. 10. The sensitivity is shown for  $j = 9$ , and for other values of  $j$  the shape and values are the same as shown in Fig. 10 for  $j = 9$ . It would require an extremely high degree of confounding to violate sequential ignorability in our sample because the estimates of  $ACME(j)$  will turn 0 only when  $\rho$  is 0.6, which is a very high value (given the nature of our sample).

**Figure 11: PEER INFLUENCE CHANNEL IS VALID IN GENERAL**



ATE: The average  $j$ -period ahead peer influence effect during July-Dec 2010. This curve is the same as in Fig. 8(c). ACME: The average  $j$ -period ahead peer influence effect mediated via Channel 1 (creation of new projects) during July-Dec 2010. Error bars are scaled standard deviations of  $j$ -period ahead influence estimates.

There is an alternative interpretation [38] for the sensitivity test:  $\rho^2$  is the product of unexplained  $R^2$  values of models (2.3) and (2.4). So the value of  $\rho^{j^2}$  for which  $ACME(j)$  estimates will turn 0 is 0.36, i.e.,  $ACME(j)$  estimates in Fig. 9 can be refuted if  $\rho^{j^2}$  in our sample is 0.36. Let us consider the case of  $j = 1$ . In our estimated models, the  $R^2$  values for (2.3) and (2.4) are 0.028 and 0.43 respectively.  $R^2 = 0.051$  in model (2.1) (Table 12(b), Model 2) increases to  $R^2 = 0.43$  in (2.4) when the mediating variable is introduced for analysis of the causal pathway of treatment

effect. To obtain a product value of 0.36, for instance, with a confounder explaining about 60% of the unexplained  $R^2$  of 0.972 in (2.3), it still needs to explain about 0.62 (108%) in (2.4) which is well above the maximum  $R^2$  value of 100%.

So based on above sensitivity tests, we can not claim that the sequential ignorability assumption is violated in our empirical analysis. The estimates obtained in Fig. 9 are therefore valid, and provide a causal interpretation of the mediation of peer influence in production popularity via the creation of new projects by Scratchers. Finally we present a general validity of the peer influence channel. For this we estimate  $ACME(j)$  curve for each of the last six months of 2010 (to allow a direct correspondence to Fig. 8(c)). The average values of the  $ATE(j)$  curves and  $ACME(j)$  curves during this period are shown in Fig. 11. This plot confirms that creating new projects in future forms an important channel due to which the production popularity of Scratchers increases in future periods, in response to existing production popularity of their peers.

## 2.5 Discussion

We analyzed peer influence in the feature-rich Scratch community under a quasi-experimental setting. Our method accounts for peer influence after controlling for various other mechanisms that can lead to clustering of behaviour in friendship network, i.e., Scratchers and their peers having similar or dissimilar behaviours. We saw how peer influence affects behaviours (creation of new projects, consuming similar projects) and outcomes (production popularity) in the collective learning environment, Scratch. We found a persistent effect of peers’ production popularity on Scratchers’ future production popularity, and that a large proportion of this effect is mediated by Scratchers’ decision to create new projects. We also saw that users who specialize in creating “remixed” projects are susceptible to peers’ production popularity more in the short run than in the long run. For consumption behaviour, we found that Scratchers are not influenced by their peers. In particular, we saw that the tendency of users to consume projects from specific sources can not be attributed due to in-

fluence from their peers. Although the primary aim of this study has been to understand the role of peer influence in the production and consumption of projects on the Scratch platform, the study has also provided several methodological insights that can inform future work. In retrospect, we believe our results are surprising, especially because production popularity of peers (resulting from the likes of several other users) influences self-decision making to create new projects which would not have been created in absence of such popularity of peers' projects. New projects created due to peer influence attain higher popularity in future compared to projects which are created by users who do not follow "popular friends". (The Scratch platform that provides such a collective sharing and exchange of ideas on a digital platform is therefore valuable because peer influence may not exist if users were to create projects without such a wide exposure to others' projects.) Our study contributes to the literature on production and consumption behaviour [39, 40, 41, 42, 43] (including peer influence [44, 45]) in various knowledge sharing platforms [46, 47, 48, 5]. We believe our results are relevant for a broad audience including network science researchers and practitioners and designers of future educational platforms [49, 50, 51, 52].

### 2.5.1 Limitations stemming from data

It would be definitely better if we could have data on several more factors to ensure a higher reduction in bias. (Unfortunately, such fine details are not available in the dataset we have.) So here we discuss some potential limitations of our analysis from a context of the data available to conduct the analysis. *First* is the representativeness of the sample, i.e., we only analyze users who actually joined the platform and interacted in some ways depending on how well they liked the platform under situations existing at the time of their joining. Since the study includes users coming from various countries, we expect that the results hold true in general. *Second*, comparing changes in future behaviour across treated and control groups does not eliminate bias arising out of unobserved confounders that are heterogeneous across the groups – for instance, it may be the



case that the general ability of users to operate on the platform might be more in the treated group, and so they are able to find better peers and also can produce more projects. Such confounding effects arising out of ability are however low in our study because we found a significant persistence effect when Scratchers were matched on all their personal characteristics (Table 11(a)) – since this includes various attributes, we expect it also captures the ability of Scratchers which is unobserved. Also, we do not have a priori reasons or information to predict why unobserved variables (like ability) might be distributed differently across experimental and control group, especially in presence of a well balanced matched sample. *Third*, there is an implicit assumption that a Scratcher follows another user in order to be informed about the user’s future activities. However, Scratchers happen to follow users for other unobserved reasons as well. Such reasons include help received in a project, social contact in real life, received friendly comment on a project, and joining a particular gallery [53]. Although we account for selection mechanisms in network formation by including peers’ observed confounding attributes, our estimates do not control for unobserved selection processes as mentioned above. *Fourth*, we do not have additional data (e.g., survey data [54]) to know exactly the decision making process of Scratchers. Our assumption about a Markov nature of decision making was motivated by an intuition of decision making in large social networks in general (which has also been used in several studies concerning social networks [14, 55, 56]). Although additional data might have been helpful to validate such an assumption, we believe this assumption is not too strong. The assumption used for this study is a weak assumption in the sense that although it does not capture the trend of past behaviour, it captures a summary of the trend (the aggregate count) which we believe is reasonable. This is because such aggregates (over the entire history) are the only statistics available when a Scratcher browses another Scratcher’s profile before following, and when new users who join the platform see about others’ projects and activities, and decide their future activities.<sup>9</sup>

---

<sup>9</sup>A strong Markov assumption, on the other hand, would mean that future decisions are made using information (personal and peers attributes) of aggregate activities from

Given the vastness of the users and projects, and the complexity of the existence of several interactions on Scratch platform, we believe that it is reasonable to assume that the average population (mostly composed of young children) decides its future activities based on the current state of activities and not by the heterogeneous trends in the past leading up to the current state of users' network and projects characteristics. (We have previously mentioned the plausibility of such an assumption in the context of Scratch platform in Section 2.3.)

## 2.5.2 Validity and interpretation of results

We saw that exact matching produces extremely low bias in treatment assignment compared to propensity score matching [24] which, if employed as a tool for analysis, would require a more careful inferential analysis [57, 58, 59]<sup>10</sup>, especially in presence of many features or variables describing the platform (users, projects, users network, various interactions). We believe that the peer influence estimates have low degree of bias. Although this comes at a cost of more than 50% reduction in sample size (refer Tables 10 and 12), we expect that there should not be an abrupt loss of generality of the results when speaking about the entire population of Scratchers. This is because matching procedure has been performed at different time periods to produce persistence curves as shown in Fig. 8 (c) and Fig. 11. Despite the fact that the users who are dropped out of analysis due to constraints of exact matching are random and not in control of the researchers, these curves are quite smooth and have similar patterns and estimated values as in Fig. 8 (b) and Fig. 9 (B) respectively. Hence we believe the results are true for the entire population of Scratchers at large.

We believe the description of exact matching in Section 2.3 is sufficient for the purpose of our analysis since we achieved both a reasonable

---

the current month alone. However a weaker assumption allows for future decision to be based not only on the current month's statistics, but also on all previous months' activities summarized as an aggregated counts.

<sup>10</sup>Although a balance of propensity scores is necessary for removing selection bias [24], it is not sufficient – the confounders should also be balanced across treated and control groups.

balance of covariates (and much better than propensity score matching), as shown in Table 11, and a reasonable number of observations for statistical estimation and hypothesis tests of significance of peer influence effects, as shown in Table 12. However, we would like to provide additional details for interested readers to explain better the role of using exact matching in causal estimation of the peer influence parameters [35]. We need to note that matching is not an estimation method. The essence of estimation strategy used in this study is to compare future changes of Scratchers with peers having higher degree of behaviour (experimental group) to those Scratchers whose peers have lower degree of behaviour (control group) in a way which can be argued to be of experimental standards even if we only have observational data. To achieve this goal, exact matching has been used as a data preprocessing step and has helped us in several ways. First, it helped us to create the experimental group when the treatment (peers' variable) was continuous. This was achieved by dichotomizing the treatment by thresholds, a recommended practice [35], which were also shown in Fig. 7 to not influence the nature of our conclusions. Second, due to its intrinsic property, the exact matching helped us to create groups to mimic experimental standards by achieving balance of covariates. Since exact matching on large number of variables generates loss in data, we always ensured to focus on matching the most significant confounders first. Understanding which variables might be important confounders in the data is done by analysing columns (3) and (4) in Table 10 to understand selection into treatment. Unmatched or variables with poor balance in matching were always included as a part of estimation method during regression analysis. Third, it helped us to remove model dependence from our analysis [35], i.e., the peer influence estimates are not extremely dependent on the variables chosen for regression analysis. This is the reason of presenting two different models in Fig. 6 - the estimates are stable. However we use the dataset matched on both personal and network variables for later analysis because this has the best balance of covariates. Lastly, we would like to mention that "exact matching" does not mean having "exactly similar observations" in the control group for each observation in the experimental group. (In fact,

such a situation would be impossible, especially in the space of high dimensional features.) The resulting dataset produced after exact matching has the property that all variables representing personal or peers' characteristics of an ego have the same observed empirical distribution across experimental and control groups. This is in line with the true meaning of exact matching and the requirements for avoiding selection bias on observable variables [35]. We would encourage readers to interpret the estimated values of peer influence as upper bounds, to allow for decrease in estimates due to potential (unknown) unobservables which might be distributed unevenly across experimental and control groups.

Identification of peer influence in our empirical strategy solely relies on the non-existence of confounding latent variables. As mentioned above in discussing our limitations stemming from data availability, we do not claim absence of unobserved variables.<sup>11</sup> If Scratchers and their peers exhibit homophily on a latent covariate and this covariate is correlated both with peers' behaviour (e.g., high or low degree of peers' production popularity) and ego's future change in behaviour, only then such a variable is a confounder, conceptually. This is because only in this case one can claim that the change in future behaviour of ego was driven by common shocks from the latent variable and not due to peers' behaviour. We believe that such confounding variables are least likely in our analysis. *First*, covariates included in analysis, as shown in Table 8 reflect individual preferences for producing projects (e.g., total projects and remixes), individual preferences for consuming projects (e.g., favorites, comments), collective preference of platform users to consume an individual's projects (e.g., love-its, downloads, favorites, comments), attributes reflecting general statistics of platform usage (e.g., age, activity), peers' attributes of all individual properties mentioned above, and characteristics of local network. These variables already reflect a wide range of individual preferences for behaviour on the platform to create projects and build peers network. Activity statistics, which is an important rep-

---

<sup>11</sup>Peer influence estimates can become less biased if we could have data on further specific details that reflect individual preferences. However, we believe that the current dataset already has details of a wide range of features that summarize well the activities on the platform.

resentation of how a user understands the platform details, are well balanced for individuals and their peers. *Second*, as shown in and Fig. 10, the sensitivity test demands very high values of  $\rho$  (0.6) for violation of sequential ignorability assumption required for mediation analysis conducted in Section 2.4. Given that the variables are well balanced (Table 11), we believe that the likelihood of existence of confounding variables in our analysis, enough to violate sequential ignorability assumption, is very remote. Since variables that can violate sequential ignorability assumption are also the ones that pose threat to peer influence estimates, we believe that latent variables which are of confounding nature are a least likely case in our study.

### 2.5.3 Behaviours analysed in this study

While there are several production and consumption behaviours that may be analysed on Scratch platform, our choices of production popularity and consumption specificity were guided by the following reasons. (We encourage investigation of other behaviours in Scratch, and also in other platforms, in future.) *First*, we wanted to ensure that the behaviours we investigate are plausibly widely known in the Scratch community. For example, a Scratcher knows about his peers' production popularity (e.g., when somebody loves one of his peers' projects) and consumption patterns (e.g., when one of his peers consumes project by favoriting it) through activity feeds. Although a user may not assimilate everything that shows up on activity feeds in real time, we believe that a general knowledge of repetitive behaviour of such peers' activities over a certain duration of time might influence the user to adopt similar behaviour. *Second*, it seems that popularity is a factor that affects social behaviour in general (outside Scratch) [60]. Popularity may be indicative to users who are aware about the activities on the Scratch platform about the popularity-to-quality ratio, i.e., popularity of a project might be indicative of the project content (e.g., codes, creativity, etc.) and hence other users might be interested to learn such things. In this sense, project popularity on the Scratch platform may be seen as a form of col-

lective assessment of the project and thus an increase in a user's production popularity may correspond to an improvement in his/her (unobserved) ability to create better projects. So production popularity may not be as irrational, as a factor to generate influence on self-decisions, as it may otherwise seem to an individual not using the platform. (In retrospect, we indeed find users' behaviour being influenced by peers' production popularity.) *Third*, while peers' projects may influence a user's behaviour, the user might be influenced to a certain extent to consume projects similar to his peers. While tracking each project for each peer is an extremely unlikely situation, we believe that a consumption influence might exist if a Scratcher observes that most of his users tend to consume a "certain group" of projects. While there may exist several ways to identify such groups of projects, we believe our strategy is feasible on a large scale [61] and also conveys important meaning. We categorized consumption baskets/groups in an intuitive fashion (analogous to various products in a supermarket): projects that are consumed together by most users were placed in one category. (In doing so, all projects have been included in one of the communities and there is no loss of observations.) Later we found from our analysis that users do not tend to be influenced if their peers have high specificity for such a consumption source/category. In retrospect, we investigated and have clearly stated that such communities do not correspond to themes or topics of projects, and neither the choice of network algorithm to detect communities affect our findings. We also believe in retrospection that the inability to differentiate such communities by a particular attribute can be a potential reason why no peer influence exists for consumption patterns. In fact, if the consumptions baskets are largely similar to each other, the reasons for switching to peers' consumption patterns is minimal. (We believe that specificity of consumption of projects is potentially a result of the local network to which a user gets associated to during his/her joining to the platform and formation of initial local friendship network.) In any case, our choice of analysis of consumption behaviour was largely guided by general intuition rather than alignment with forward-looking results. *Fourth*, a user could have hundreds of peers whom he follows but

it seems implausible to be influenced by each of them in a heterogeneous and meaningful way. Therefore we used the aggregate measurements of each attribute as a potential source of influence. Later, as shown in Table 1, we found that actually there is no effect of variance of peers' popularity on how users are influenced. In other words, the influence from production popularity is largely an aggregate effect from the popularity of all projects from all peers of a user and not an effect arising from specific peers.

### 2.5.4 Some topics for further investigation

We discuss several studies that may be done in the Scratch platform and other digital platforms. *First* is to understand the aspect of "learning" more precisely. In this study, we saw that Scratchers decide to create projects (new, remixed) in future due to peers' influence; some of the new projects may be totally copied versions without being assigned as remixed projects. So due to the nature of the available data, we can not be precise about how much Scratchers actually 'learn' during the process of creating new projects. *Second*, analysis may be done using other assumptions about users' behaviour which can lead to new insights. Such assumptions can be identified from surveys, or observed behaviours on other platforms. *Third*, for peer influence analysis, especially with many attributes as in this study, ways to reduce dimensionality of the attribute space and their effects on the bias of influence estimates may be conducted. *Fourth*, new studies may be done to better understand differences in peer influence in digital contexts (as in Scratch) and physical contexts (as in classrooms). A key difference between online platforms and classrooms in formal educational systems is that, in most cases, children do not choose to go to schools whereas they usually choose whether to join a platform. Peer influence investigation in physical settings shows mixed evidences [4, 5, 62] of positive and negative influences. New studies can therefore bring clarity into subtle nuances of how children in educational environments are influenced by their peers.

## Chapter 3

# Polarization in Twitter Social Media, a Social Media Platform

### 3.1 Introduction

*Polarization*<sup>1</sup> of beliefs is about the existence of opposing beliefs within large sections of the society [64, 65, 66, 67]. In situations, like climate change action, where unanimous belief can drive the required collective steps, polarization can be a hurdle and may lead to socially undesired actions [68, 69, 70, 71]. Public understanding of climate change has remained polarized [72, 73, 74]. Polarization of beliefs can be affected not only by the nature of users' tendencies like homophily in communication about the reality of climate change, but also by certain nature of the information itself. Credibility of information is one of such properties of information. We discuss these determinants of polarization below.

*Homophily* in communication can affect the polarization of beliefs. Homophily in communication is a key tendency of users in informa-

---

<sup>1</sup>This chapter is based on the following published work:  
Samantray, A., Pin, P. (2019) "Credibility of climate change denial in social media." *Palgrave Communications* 5, 127 [63].



tion propagation media like Twitter and other platforms: people tend to communicate with others who hold similar beliefs [75, 76]. Although homophily can facilitate the flow of information [77], it can lead to polarization [78] as well and decrease the general ability of the society to learn the truth [79, 80]. Communicating only with people who share similar ideology or opinion restricts beliefs and prevents learning the truth [81, 82, 83]. For example, homophily among liberals and conservatives in political blogs links [84], i.e., the tendency of liberal and conservative blogs to link primarily within their separate communities, leads to echo chambers [85, 86, 87], and hence beliefs can be polarized due to such exposure to selective information [88]. Similarly, homophily in opinion exchanges in social media (e.g., via publicly visible replies and mentions in Twitter) can reinforce beliefs within various sections of the network due to selective exposure. Previously echo chambers have been observed for climate change discussions on social media [89, 70]. Such patterns when repeat themselves in various parts of the network can lead to polarization of beliefs.

*Credibility* of information is also an important factor to create polarization, especially in online media where usually there exists information from several sources [90] that propagate controversial beliefs. Credibility of information is the precision of information, it signifies how certain the information is and helps to assign a certain level of trust to the information [91, 92]. The credibility of information that propagates in a social network is a critical factor that can shape the beliefs: if incoming information from a person's social network carries no credibility, then it is less likely to be incorporated in to the belief of the person [93, 94]. Hence, negative consequences that may arise due to the spread of fake information in social networks depend on information credibility, thereby making it a factor that can induce polarization. The importance of precision of misinforming signals has also been highlighted recently by Allcott and Gentzkow [95]. Credibility is a dimension of information which is independent of veracity of the information. For instance, a particular (unintended) fake story can be more credible if the information source is highly reputable (and hence the fake story inherits the credibility of

the source) compared to the case when the same story comes from a less credible information source. It is the former case that has the potential to change beliefs in the society and change the existing levels of polarization concerning the truth of its story. Credibility of information in social media, generally speaking, may be ascribed due to several factors including reputation of information source, number of verified facts cited along with the information, mentioning opinion leaders and others [96, 97]. Lastly, credibility of a communicated information is independent from the preciseness with which the information source believes the information. For example, a certain propaganda house which believes strongly about a story need not be able to spread such beliefs in the society through various communications because such communications may lack the degree of credibility required to generate substantial change in beliefs in society.

Twitter has become a modern platform for news dissemination and opinion exchanges, and is widely adopted by many users worldwide. We believe discussions and information propagation that happens on such a platform has potential to shape beliefs at a large scale. Important topics like climate change are also discussed on such social media platforms. With fake information on many topics becoming prevalent on widely adopted social media platform like Twitter, it is probably not surprising that there are several tweets both in favour and against the statement that climate change is a real concern. Fake information regarding such an important issue as climate change can pose a collective hurdle for the society at large if such information becomes highly credible among users of the platform. In this study, we infer the credibility of anti-climate change opinions on Twitter using (i) an empirical analysis that investigates the effect of homophily in communication patterns on the polarization of beliefs, and (ii) the predictions of a model of polarization of beliefs that jointly accounts for the roles of information credibility and homophily in communication. For the empirical analysis, we use tweets about climate change topic during 2007-2017. We rely on opinion exchanges among climate change believers and sceptics made via mentioning (and replying) others in tweets as the communication pattern.

(Replies in Twitter tag the user names at the beginning of tweets, while user mentions can be made anywhere within the text of tweets.) This is because retweets do not contain new opinions and are mere repetitions or broadcast of opinions expressed in original tweets, whereas mentions contain explicit referencing of other users and hence convey exchange of opinions targeted towards users being mentioned.

## **3.2 Methods**

### **3.2.1 Data**

We use tweets from the online social network site Twitter. Tweets are the messages posted by users on the platform. Tweets can contain hashtags, mentions to other users, external links in addition to text. A tweet posted by a user can be retweeted, replied and liked by other users. The tweets from 2007 to 2017 were collected based on a search filter that each tweet contains at least one of the following words: ‘climate change’, ‘#climate-change’, ‘global warming’, ‘#globalwarming’. The search was performed via Tweeter’s public advanced search page. This collection of tweets does not violate any ethical standards and consists of only publicly available data. The data is complete in the sense that it contains all original tweets that fulfil the search criteria. Retweets of original tweets are not included in the data. Each tweet however contains the retweet statistic, i.e., how many times it has been retweeted. There are a total of 14,353,859 unique tweets (without counting retweets values) and 3,595,205 unique users in the dataset.

### **3.2.2 Measuring Sentiment & Opinion**

The sentiment of each tweet is computed using VADER [98] model which is designed to conduct sentiment classification specifically on short texts like tweets. For each tweet, a score is obtained on the scale -1 (most negative) to 1 (most positive). A message with a positive sentiment is usually in favour of the (intended) object in the message. However, the same

message may be communicated in a way that contains negative sentiment. Therefore sentiment and opinion contained in a message are two different characteristics of the message. This distinction between sentiment and stance has been clarified by previous studies on stance detection of tweets [99, 100]. From various data used in these previous studies, we use a subset that is used to annotate whether the tweet is in favor or against the statement “climate change is a real concern”. This subset contains manual annotation of whether a tweet is in favor or against the statement. We use this data as the training data to predict the opinion expressed in the texts of tweets in our sample.

The tweets were coded as numerical features using the TF-IDF (term frequency-inverse document frequency) representation and the prediction on the dataset for this study was done using the support vector classifier, which had the highest predictive power among other classifiers (logistic regression, decision tree). Each user’s tweets were first classified into one of the following categories: in favour of the statement, against the statement, no opinion. (Hence, in the first stage, opinion is assigned to each tweet.) Next, each user was classified into one of the above categories based on the category that contains the maximum number of tweets from the user. Since a user on Twitter usually has a belief about climate change (in general), the classification algorithm has managed to predict each users’s tweet into a single category only. Only 42 users were classified as having no opinion, and were dropped from the dataset. Most users had tweets in one category only: either in favor, or against the statement. This validates the performance of the classifier at the user level even if the prediction accuracy (at the tweets level) is about 72%.

### **3.2.3 Measure of Homophily**

The notion of homophily in communication is the presence of higher interaction among people who hold the same opinion about climate change. For instance, the group of users who believe climate change is not real have a homophily in communication pattern if such users communicate

more among themselves than with users who believe climate change is real.

Communication between users can take the forms of retweeting and mentioning (includes replying) on the Twitter platform. Although retweeting, in general, suggests that users share the same ideology, mentioning can occur between users with differing ideologies (tweet wars). To what extent do mentioning patterns in tweets reveal communication among users having similar ideologies? Previous studies [101, 102] reveal that homophily of the aggregate network measured using retweeting and mentioning tend to be same, although mentioning is more volatile at the group (sub network with a particular ideology) levels. Hence, based on such studies, we assume that homophily in mentioning patterns reveals the homophily in retweeting (and following) patterns, for which we do not have the data. Communication therefore refers to users interacting via mentioning each other in tweets.

To measure homophily in communication among users with same belief or opinion regarding climate change, we follow the measures used in previous empirical studies on homophily [103, 104, 105]. Let  $I$  be the total number of users in a given month who either mentioned other users or received mentions from other users' tweets. We indicate belief of a user as  $b \in \{a, f\}$ , where  $a$  and  $f$  correspond to the cases where the user is against and in favour, respectively, of the climate change statement.  $a$  and  $f$  represent two types of belief  $b$ . Suppose  $I_b$  is the total number of type  $b$  users, then  $w_b = \frac{I_b}{I}$  is the fraction of type  $b$  users. Let  $v_{ib}$  be the number of type  $b$  users mentioned by user  $i$ . Then  $s_b = \frac{1}{I_b} \sum_{i \in I_b} v_{i,b}$  is the average number of same type users mentioned by type  $b$  users, and  $d_b = \frac{1}{I_b} \sum_{i \in I_b} v_{i,-b}$  is the average number of opposite type users mentioned by type  $b$  users. Using this, homophily of the group of type  $b$  users is defined as  $H_b = \frac{s_b}{s_b + d_b}$  and the homophily of the society is defined as

$$H = \sum_{b \in \{a, f\}} w_b H_b.$$

### 3.2.4 Measure of Polarization

A tweet carries an opinion (whether in favour or against climate change being real), denoted as *op*, with a certain sentiment, denoted as *s*. (We encode *op* as 1 if the message in the tweet is against the climate change statement and as  $-1$  if the message is in favour of climate change.) Irrespective of the opinion, the sentiment can be positive or negative depending on the way the message is communicated. So we use an emotion-adjusted measure of belief called EAB that combines these two aspects: expressed opinion in the message, and emotional content in the message. For a given tweet with attributes *op* and *s*, EAB is defined as  $op \cdot |s|$ , the product of opinion and absolute value of sentiment.

A large number of tweets have neutral or close to neutral sentiment, thereby increasing the mass around 0 and decreasing the intensity of bi-modality in the distribution of EAB. (The presence of high volume of neutral tweets is not uncommon, for example see Kušen and Strembeck [106].) In this sense, the polarization indicator used during the mathematical analysis in opinion updating model (see Appendix B.3) cannot be carried to the empirical setting directly. We therefore need another measure of polarization that is sensitive to the properties (e.g., kurtosis and skewness) of a distribution and carries the intuition of bi-modality approach (i.e., the statistic should be sensitive to the degree of bi-modality of a distribution). To calculate the degree of polarization involved in the EAB distribution, we use the measure of ideological divergence provided by Lelkes [107], which characterizes the level of polarization based on bi-modality of the distribution by being sensitive to kurtosis and skewness [108, 109]. Using this measure, the polarization of EAB is defined as

$$\frac{s^2 + 1}{k + 3 \frac{(n-1)^2}{(n-2)(n-3)}},$$

where *s* and *k* are the skewness and excess kurtosis of the EAB distribution, and *n* is the sample size. The values of 1 and 0 correspond to cases when the EAB distribution is perfectly bimodal and perfectly unimodal respectively. A value greater than 0.56 is categorized as bimodal.

Although this threshold may not be reached precisely, approaching this value is considered coming closer to polarization [107].

The measure of polarization above is calculated at the levels of tweets and is a stricter measure than when calculated at the level of users. This is because a user might be slightly heterogeneous in his tweets but aggregating all tweets would reveal one exact ideology. For example, when a climate change sceptic user writes a tweet, it is most likely to be against the statement that climate change is real. However a few tweets of his might fall in favour of the statement. In this sense, polarization is easier to obtain at an user level than at the content (tweets) level.

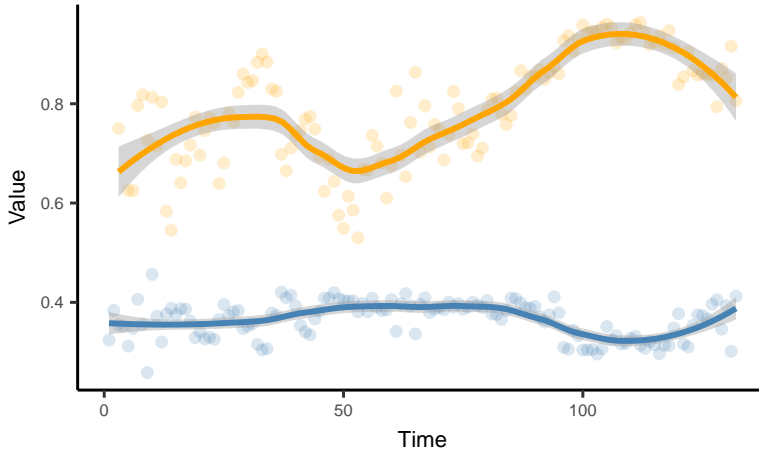
It is easy to see that the polarization measure describes a different phenomenon than the homophily measure. While homophily measure is constructed using users involved in a mention activity in a given month (micro level), polarization is measured using all the EAB of all tweets in the month (macro level). This is in line with Esteban and Ray [110], according to whom any reasonable measure of polarization must be global in nature.

## **3.3 Results**

### **3.3.1 Negative effect of Homophily on Polarization**

The monthly time series of polarization of beliefs and homophily in communication during 2007-2017 are shown in Figure 12. (Measures of homophily and polarization are described in Methods.) It appears that the evolutions of the two curves are not independent, and polarization tends to decrease at times when homophily increases. Augmented Dickey-Fuller test and Phillips-Ouliaris test suggest that polarization and homophily are cointegrated, i.e., they are bound in a long term equilibrium relationship such that they mean-revert whenever there are deviations away from the equilibrium. (Appendix B.1 contains statistical details of tests of cointegration.) Hence it becomes natural to model polarization and homophily jointly in a vector error correction (VEC) model [111, 112]. VEC models are a special class of models derived from

**Figure 12:** POLARIZATION OF BELIEFS, HOMOPHILY IN COMMUNICATION



The figure shows monthly evolutions of polarization of beliefs about climate change (blue) and homophily in communication patterns on Twitter (orange) during 2007-2017. Time 0 corresponds to Jan 2007. A long-term relationship seems to exist: polarization decreases whenever homophily increases.

vector autoregression (VAR) models with an additional term, called error correction term, to account for the cointegration. VEC model allows to study the lagged effects of homophily on polarization and also the lagged effects of polarization on homophily. Table 16 in Appendix B.2 contains details of the parameter estimates of VEC model: the lagged homophily covariates negatively affect polarization in the long term with high statistical significance and the lagged effects of polarization on homophily are not significant. These estimates however do not form a conclusion about the direction of causality between homophily and polarization. For assessing causality, we perform Granger causality test [113] with modifications, as suggested by Toda and Yamamoto [114], to adapt to the non-stationary nature of both time series; the results are shown in Table 2. (Further statistical details regarding these tests are available in Appendix B.2.1.) The conclusion that emerges is that only homophily



Granger-causes polarization with a negative effect, and the causality in the other direction is absent, i.e., polarization does not Granger cause homophily. This empirical result that homophily negatively affects polarization is counter-intuitive to the discussions made previously about the nature of homophily and polarization.

**Table 2:** GRANGER-CAUSALITY TESTS

Null Hypothesis	Test statistic
Homophily does not Granger-cause polarization	12.3***
Polarization does not Granger cause homophily	4.0

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ . Based on the test statistics, only the first null hypothesis is rejected.

### 3.3.2 Joint Effect of Homophily and Credibility

We study a simple model of emergence of polarization in social networks that highlights the joint effect of credibility of propagating information and homophily in communication of information to affect polarization. We model polarization as bi-modality of the population’s distribution of beliefs and it captures the following basic features, as laid down by Esteban and Ray [110], which are necessary for a distribution to be considered polarized: there must be a high degree of homogeneity within each group (whose agents hold same belief), a high degree of heterogeneity across groups, and a small number of significantly sized groups. If the groups are of insignificant size (e.g., isolated individuals), they do not contribute to polarization. The beliefs can be emotion-adjusted as well, as is done in the empirical analysis (see Methods), to include the emotional content of the belief; this does not change the nature of predictions of our model.

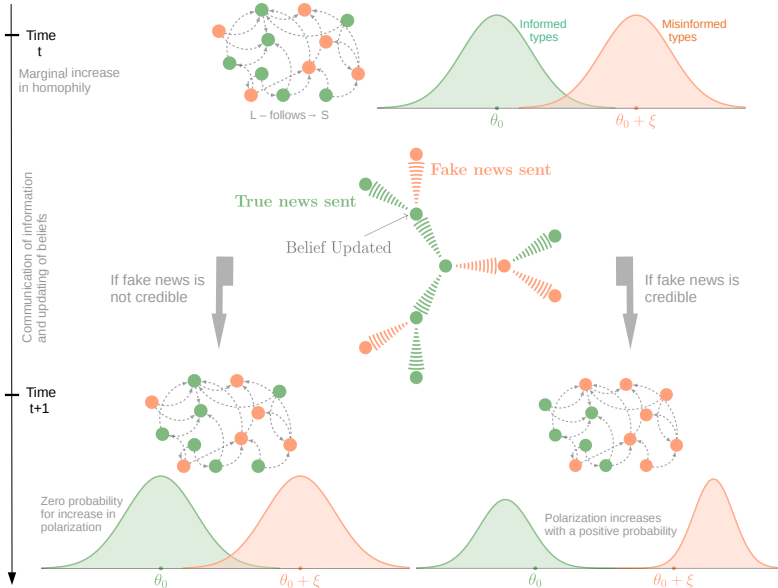
In our model, we consider a social network with each agent receiving information about a topic (e.g., the reality of climate change) from the same number,  $k$ , of speakers (for simplicity). Speakers of an agent are other agents in the network from whom he receives information. An

unobserved fundamental  $\theta$ , with values in the real line, describes the accuracy of the beliefs: an agent with a higher value is more distant from the truth. At a given time  $t$ , we assume there are two types of agents -  $\theta_l, \theta_r$  - informed and misinformed, and the social network has a homophily coefficient of  $\frac{h}{k}$  with respect to this property:  $h$  speakers of an agent of type  $\theta_i, i \in \{l, r\}$ , are of the same  $\theta_i$  type. We model the prior beliefs of informed ( $\theta_l$ ) and misinformed ( $\theta_r$ ) agents as distributions of  $\theta$  over the real line to allow for heterogeneous prior beliefs at time  $t$ . We assume that the prior beliefs of informed and misinformed agents derive from distributions  $\mathcal{N}(\theta_0, \delta_l^{-1})$  and  $\mathcal{N}(\theta_0 + \xi, \delta_r^{-1})$  respectively with a strictly positive value for  $\xi$ , in order to ensure that the misinformed type are, on average, farther away from truth compared to the informed type. The belief of the entire population of the social network at time  $t$ , a density-weighted average of the above distributions, is considered to be polarized if it has two modes (peaks in the distribution).

We assume each agent communicates a particular realization of his belief using a message (e.g., via a tweet) during the period between  $t$  and a future time  $t + 1$ , and all agents update their prior beliefs at  $t + 1$  after incorporating beliefs about the fundamental expressed in their speakers' messages. Mathematically, a message from a speaker is modelled as an independent noisy signal about the speaker's realized belief, with the degree of the noise (or uncertainty) being conditional on the type of speaker. Since credibility of a communicated message is the precision of belief expressed in the message, in our model, true and fake information about the reality, arising from informed and misinformed speakers respectively, propagate with different credibilities in the social network. (We denote the credibilities or precisions of true and fake information as  $\beta_l$  and  $\beta_r$  respectively.) We assume that the messages from informed agents carry a positive credibility. (In our context of the reality about climate change, this is a reasonable assumption.) It is noteworthy to mention that for a speaker of a given type  $\theta_i, i \in \{l, r\}$ ,  $\beta_i$  is conceptually different from  $\delta_i$ : while  $\delta_i$  characterizes the probability with which a particular random realization will be selected as the belief for the speaker,  $\beta_i$  characterizes how precisely the realized belief is communicated in a

message. In a Bayesian update of beliefs, how a communicated message affects the beliefs of listeners of the message depends on the belief expressed in the communicated message, the precision (or credibility) of the message, prior beliefs held by individual listeners and the precision of such prior beliefs of listeners.

**Figure 13: BELIEF UPDATING MODEL**



The figure illustrates increase in polarization at time  $t + 1$  due to increase in homophily (based on beliefs, e.g., reality of climate change) at time  $t$ . There is zero probability for such an increase if fake information that propagates in the social network has zero credibility.

We now present the findings of some analyses conducted using the model discussed above. (Mathematical proofs of the results presented below are available in Appendix B.3.) We find that in a random communication network with agents being sufficiently uncertain about the prior beliefs they hold, polarization can not arise at time  $t + 1$ , after agents have incorporated their speakers' beliefs. This means that, irrespective of the

precisions with which true and fake information diffuse in the social network, they do not play a role in the emergence of polarization without the support of non-random topology of the social network and agents' prior beliefs. We also find that if marginal increase in homophily at time  $t$  does not increase the probability of polarization at  $t + 1$ , it can happen only in the case when the fake signals that propagate in the social network do not have a minimal level of credibility and carry zero precision. A visual illustration of this result is shown in Figure 13. We believe this provides a potential reason for our previous observation that, in the climate change discussions on Twitter, polarization of beliefs does not increase when homophily in communication patterns increases.

### 3.4 Discussion

The online social network Twitter has remained an important media for rapid spread of opinions. We studied the opinions expressed in Twitter during 2007-2017 regarding the reality of climate change. Below, we briefly discuss our main findings and their limitations, and suggest directions for future research.

The analysis over a long time period provides insight about the direction of potential causality between homophily and polarization which would otherwise not have been possible in cross-sectional observations. In social networks, polarization of beliefs (existence of large groups of people with opposing beliefs) and homophily in communication (communication among people having same beliefs) tend to be highly correlated. One potential mechanism for such a correlation is that increase in homophily can reinforce individual beliefs leading to the creation of echo chambers and hence increase polarization. Another mechanism for such a correlation is that increase in polarization increases segregation of a society into different beliefs thereby acting as a natural source to increase the probability of like-minded communication (homophily). In the case of Twittersphere of climate change conversations, we performed Granger causality tests on the evolutions of homophily and polarization, and found that only homophily Granger-causes polarization and

not vice-versa, i.e., increasing polarization does not lead to increase in like-minded communication. It is important to note that Granger causation of homophily on polarization does not fully establish its causality. Granger causality is a concept about precedence of the cause (homophily) before the effect (polarization). It says that the evolution of homophily significantly helps to predict evolution of polarization in future. Existence of Granger causality therefore does not exclude the possibility of an unobserved variable to drive the evolution of both homophily and polarization. Although the presence of such an unobserved variable is highly unlikely within the framework of our model and the variables used for empirical analysis, we encourage future research to build upon the above results to improve the causal nature of the effect of homophily on polarization - this would require analysing scenarios where homophily can be argued to be exogenous so that we are better ensured about effects not being driven by unobserved variables. (Such causal relationship may be investigated for topics other than climate change as well.)

We also found that the effect of homophily on polarization is negative, i.e., increase in homophily in communication leads to decrease in polarization of beliefs in future. This is counter-intuitive because we would expect an increase in polarization when homophily increases. Increasing homophily leads to situations where people are exposed to a narrow set of beliefs [82, 83] that conforms to their existing beliefs. When such homophily in communication happens in two large sections of the society with differing beliefs, it enhances polarization [110].

Polarization of beliefs can be affected not only by homophily in communication among people, but also by the credibility of information that propagates on Twitter. In this study, we investigated whether credibility of information source plays a role to increase polarization. We studied the ‘credibility’ factor because this has received less attention in the literature and is a very intuitive factor. It is intuitive since information from a source which is not credible is naturally least likely to affect or change the belief of an individual. Credibility is a dimension of information independent of the veracity of information. For example, let us consider

a certain information (either true or fake) that comes from two different sources at the same time: one source is a Twitter account of a person who is not very well known, and another is the Twitter account of BBC News (as instance) which has a much larger credibility in the society. In this case the news sent by BBC News has a higher probability to influence the beliefs of people at a large scale. More so, if the news happens to be (intentionally or unintentionally) fake, it can possibly affect the beliefs of a certain section of the society in a wrong direction (since the news is fake) thereby increasing polarization in the society because a new section of the society emerges with beliefs much different from what the entire population believed previously.

We modelled these two determinants of polarization discussed above, homophily in communication and credibility of information, jointly in a belief updating model where agents in a networked society receive true and fake information from their neighbours. We modelled the credibility of each type of information using the precision or certainty of the information. The model predicts that marginal increase in homophily always leads to increase in polarization except for the only case when fake information has no credibility. In the case when fake information is not credible, the model predicts a negative effect of homophily on polarization. (This description is illustrated in Figure 13.) Since we observed a negative effect of homophily on polarization in the empirical analysis of climate change discussions on the Twitter platform, we conclude that tweets expressing anti-climate change beliefs are largely not credible to the broader society.

Our results disentangle the *presence* of fake information on social media from its potential *negative effect* on the society. The spread of fake information (either as misinformation or disinformation), as has been prevalent during recent years in online social media [115, 116], has the potential to pose harm to the society by polarizing the beliefs of people [115, 116, 117]; for instance, it can influence political election outcomes [118]. However, does it mean that fake information always has a negative effect on the society? Based on our results of this study, we can say that it is not always the case. In the case of reality about climate change,

although there are many climate-sceptic messages, we showed that such messages are largely not credible and hence polarization (a negative consequence for society) does not increase when users on the platform communicate among themselves. Hence, in general, we believe that the presence of fake information (on various topics including climate change) in social media and the web is not a conclusive evidence of its negative effect on the society at large. This also reaffirms the fact that the negative effects of echo chambers in social networks, which are known to arise due to high levels of homophily in communication, might be overstated [119, 64, 87, 120].

We now discuss some assumptions about agents in our model, and how alternative assumptions about human behaviour (in real world) may also explain how increase in homophily can lead to decrease in polarization in future. Agents in the model are rational [81] and purely rely on previously held beliefs and beliefs expressed by neighbours in their immediate social networks to arrive at new beliefs. Such assumption of Bayesian updating of beliefs dictates that when homophily increases (i.e., agents' exposure to non-conforming beliefs decreases, or in other words, cross-attitudinal interactions decreases), the ability of misinformed agents to move closer to truth decreases. Human behaviour, in general, is highly heterogeneous and there may exist different ways in which people update beliefs in real world. Let us consider a different situation where the utility of agents depends not only on the belief of own type but also on the (opposite) belief held by the other type. In particular consider a situation where people update beliefs in the following manner - when people are exposed to cross ideologies, instead of gaining higher utility by incorporating and averaging it with own ideologies, they gain higher utility by the fact that they know something (what they think is supposed to be known) and the opposite party is misguided and carries the wrong belief. In such a case, if homophily increases, cross-attitudinal interactions decrease and the misinformed agents are less convinced about their belief because their exposure to informed type (opposite type) agents decreases. This leads to dilution of ones's held beliefs and creates a decrease in polarization. Although such

a behavioural assumption can lead to an alternative way to explain negative effect of homophily on polarization, we believe that such assumption may not scale to a wider population. (Especially in the case of reality about climate change, we believe that informed people would also want ill-informed sections of the society to know about the reality as they do.) In this context, we believe that by imposing a Bayesian updating criteria we have studied the expected outcomes in a baseline model, and is (ideally) representative of the behaviour of a larger population. We consider the model we analysed in this study to be parsimonious enough to be able to explain the negative effect of homophily on polarization by highlighting the role of information credibility. Future research, depending on the nature of investigation, may find our analysis as a useful guide.

Individual rumours are being debunked in due time by fact checking measures [121, 122], however, public perception about some issues like human induced (anthropogenic) climate change have remained controversial for a long time [123]. Such controversy appears to persist even after several investigations made by the scientific community. Various factors like exposure to different kind of information on the media [124, 125], politicization of climate change [126], and exposure to fake information on social media platforms are potential reasons that can contribute to polarization of beliefs about the reality of climate change. In this study, we showed that social media messages on platform like Twitter is not a potential concern because although there are many sources propagating fake beliefs regarding climate change, the collective credibility of such sources is negligible. Future research may contribute to improve our understanding about how public perceives the reality about climate change. It is better that the society is not polarized on beliefs about the reality of climate change, so that timely environmental policies can be implemented with least public resistance. Since credibility of major fake information sources about climate change can polarize beliefs on a large scale, it is important to assess credibility of such sources and their potential negative effects on society. (In this context, polarization is one such negative or undesired effect.)



## Chapter 4

# Politicization in Guardian, a News Media Outlet

### 4.1 Introduction

In democracies, citizens' opinions play important role in influencing future policies, and the mass media has been known to strategically shape public perception on various issues [127]. Hence what and how media presents facts plays a critical role in shaping public opinion and policies. Exposure to news media can cause people to participate in various issues, and join national policy conversations [128]. The more media writes about a certain topic, the more important it becomes in the eyes of the public [129]. At the same time, the ways media presents news also has its effects: a negative coverage of an issue can create a negative perception of the issue in the public [129]. So the manner of communication can have an effect on the public understanding of the issue as well. A particular instance of the way of presentation, relevant for this study, is politicization of the news articles. There are several evidences of politicization of important national and international issues [130, 131].

Climate change is an important issue facing the planet with potential for vast damages [132]. Although industry regulations of activities leading to global warming is important, change in behaviour of public

through individual and group efforts and raising voice for the needed policies is equally important. However public understanding of climate change has not been unequivocal [133], due to several potential reasons. Among such reasons in shaping public perception, climate change communication by scientific experts [134, 135, 136] and media [137, 138, 139, 140] play important roles.

The sensitive issue of climate change has also been acknowledged in the literature to be politicized [141, 142, 143, 144]. The role played in it by the ways of presenting news in media [145, 146, 147, 148] can affect how public forms opinions on the issue [149, 150], understands the importance of the issue [151], participates in discussions and taking actions [152, 153, 154], and thereby shaping future policies [155, 156, 157, 158]. For example, framing a particular discussion on global warming with mentions of scientific opinions can potentially make various attributes of the discussion more important and objective than that with mentions of political figures and party identities. In fact, a recent study [159] shows that partisanship during communication, even by only the exposure to the logos of political parties, can affect social learning and increase polarization of beliefs about climate change effects. This shows high degree of susceptibility of the perception about climate change to politicization of the issue.

News articles contain numerous options to politicize climate change, let alone the images of political parties' logos. It is understandable that mentions of factual content (scientific facts) in an article about climate change can (and should) objectively influence readers' perception on the issue. However, can changes in articles reflecting *political positioning or inclination of an article about climate change* really influence the perception and response patterns of the readers?

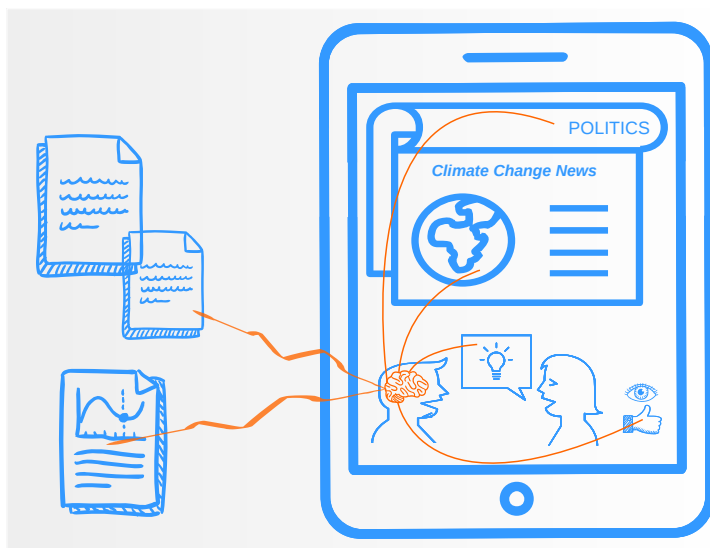
As mentioned above, previous research findings and articles on the web seem to suggest that (i) climate issue is politicized, and (ii) political framing in general by the media can influence public understanding and participation. However any concrete evidence on *the media effects of politicization of the climate change issue* over an extended time period is absent. In this study, we fill this major gap by investigating *whether* and *how* the

collective behaviour of the audience of a reputed news outlet pay attention and respond to politicization of climate change articles published by the news organization.

Using articles related to climate change from the Guardian, we infer the time-invariant nature of effects of politicization by using (the best possible) time-invariant measures of treatments for estimations, setting up quasi-experimental conditions for measuring treatment effects prior to statistical estimations, and using time fixed effects during estimations. We find a positive impact of politicization on various variables describing collective discussion/response to the articles. We investigate politicization due to positioning of articles related to climate change in the 'Politics' section of Guardian and due to mentions of politically inclined named entities within the articles. The estimates are robust to time fixed effects, authors, and potential confounders. In the case of influence from named entities mentioned within articles, we provide results for two cases of readers' behaviour depending on whether or not they might perceive the political inclinations of entities due to past association as they read a current article. In addition, we investigate the mechanisms behind the effects of politicization on user participation and other discussion characteristics. In both cases we show how politicized perception channels the effects of politicization on collective attention and engagements. The main results from these investigations described above are summarized in Figure 14.

We also conduct an investigation on the potential intent of authors to make their articles more politically oriented in order to drive collective attention and increased participation on the discussion of articles. For this, we use the risk averseness of authors as the estimate of such intent. We find that authors are risk averse instead of risk preferring, i.e., they are not likely to bet on extremely politicized content to drive attention. Although there is no issue of reverse-causality in the estimated effects of politicization on various characteristics of participation because of temporal separability of publication and discussion, this result on risk averseness says something about their intentions to politicize in future periods after seeing the effects in a certain period of time. Overall, we do

**Figure 14: COLLECTIVE DISCUSSION**



Readers of an article who join the discussion can be influenced by various characteristics of the article, past articles related to the current article, and current state of the discussion. We find that positioning an article in 'Politics' section drives higher collective attention and response to the article when compared to positioning in any other section. We find that at least 65% of users' participation in discussion due to politicization (via politically oriented textual entities in article's main text) is driven via users' collective recollection of past political contexts related to the current article. Users influenced politically in such a manner also drive at least 40% of the impact of politicization (as above) on total comments, social feedbacks, and users' engagement in the discussion.

not expect the authors, in general, to politicize climate change articles in expectation of higher collective attention to their articles.

In the following section, we discuss the data collected from Guardain and some descriptive statistics. Next, we provide details about the concepts and measures used to understand politicization in the articles. Next, we investigate the impact of politicization of various dimensions of collective attention. Next, we investigate whether authors are likely to in-

crease the mentions of politically oriented entities in articles in response to how readers respond to it.

## 4.2 Data

We use data on all articles related to climate change and the comments received in their discussion sections published in the Guardian upto September 2018. (Data was collected during the last week of the month.) The articles' texts and other metadata were retrieved from Guardian's open API service, and the comments were scrapped from the website using robots.

The metadata of articles' text consists of characteristics of the articles like publication date, author, title, section allocated to the article (Table 3), whether the article is commentable or not, and others (which are not relevant for this study). Each retrieved comment contains information on full text, its user name, time of comment, whether it is a reply to another comment, likes received on the comment, and page number. Replies are organised as a nested tree-like structure in the discussion section, as shown in Figure 15.

There are a total of 51,874 articles with text, of which only 28,570 are open for receiving comments (as decided by authors or publisher due to reasons not available in articles' metadata). The total number of comments in all the articles over the entire duration is 7,751,740 of which 7,748,575 had their texts intact at the time of data retrieval. (In other comments, the original text was removed, and replaced with a message describing so.) There are 5,046,325 replies, i.e., about 65% of the comments are replies to other comments. Rest 35% are root comments. A total of 293,330 unique users have participated in the discussion on all articles in aggregate.

As described later in Section 4.3.2, due to the nature of analysis of (joint effects of) politicization, we use certain categories of named entities as the storehouse of politicization in various articles. We keep only those articles which have at least one such entity that has appeared more than 2000 times in the entire dataset. The distribution of appearance of

named entities in the aggregate dataset contains an extremely high number of entities with almost zero frequency, compared to the remaining distribution. With this we retain 48,684 articles of which 26,737 are commentable. Among these articles open for comments, 22,171 have actually received comments.


**Table 3: SECTION SPLITS FOR CLIMATE CHANGE RELATED ARTICLES**


Entity	Meaning
Environment	14357
Opinion	6550
Politics	3700
World news	3042
Business	2734
Guardian Sustainable Business	2183
US news	1706
Australia news	1631
UK news	1331
Society	1318
Science	1219
Global development	1078
Books	925
Media	859
Education	775
Money	693
News	619
Global	527
Life and style	516
Film	384
Sport	383
Travel	350
Technology	346
Music	339
Football	294
From the Guardian	293
Stage	289
Public Leaders Network	286
Working in development	281
Art and design	259


There are a total of 128 sections in the Guardian. Each article is allocated to one particular section. The table shows a list of the top section names, ordered according to the number of articles on climate change contained in the respective sections.


Figure 15: SAMPLE DISCUSSION

Order by **Oldest** Threads **Collapsed** 1 2 3

**MasterArngelr** 22 Sep 2014 13:19  
This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

**sadhu** 22 Sep 2014 13:30 7 ↑  
All this hoolloo boolloo about climate is a good chance for people to get out of doors, meet some new people and shed their guilt feelings. As for the real things, some of us think that we have passed the point of no return. Even if there were strong measures and fines to turn the tide, which will never happen due to the power of corporations and multinationals, we still will may not be able to turn the tide. And this is not my words, go talk to the guru of Gaia, James Lovelock. He will tell you where it is out.  
< Share Report

**wakeupbomb** → **sadhu** 22 Sep 2014 15:35  
This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

**Westmorlandia** → **sadhu** 22 Sep 2014 15:40 10 ↑  
Perhaps, perhaps not. No one can be sure, of course. But sitting here and doing nothing, when it might well help to do something, seems pretty stupid. It can't hurt to do something about it - if there is an economic cost to that, it is still minimal when compared to the expected cost of climate change - both economic and human.  
< Share Report


**Wobbly** → **sadhu** 22 Sep 2014 15:50 2 ↑  
Those people out there yesterday have a chance of changing policies whereas those who do nothing have none. Two weeks ago a the redneck govt of Queensland yielded to public pressure and will not now dump dredge waste in the Great Barrier Reef. Public pressure is worth the effort.  
James who? In the last few years there have been approx. 9,000 peer reviewed scientific papers per annum on the subject. Science should be setting the agenda. In Australia, the best thing we can do is work to vote out the current climate change denying govt which is in the pocket of fossil fuel industries.  
< Share Report

Illustration of a portion of a discussion on a particular article in the Guardian.

## 4.3 Politicization of climate change

In this section we formalize the notion of politicization that we shall use in our empirical analysis later. In doing so, our goal is to use intuitive measures that reflect human behaviour, and that can be applied to large scale datasets and other contexts.

Politicization of a topic by a news organization is a broad notion that encompasses several aspects of how authors write articles for the topic over a certain time period. It encompasses the ways of presentation for promoting the topic and the ways of presentation for informing events

and happenings related to the topic in order to create a desired influence in the readers. Such manners of presentation have been discussed widely in media communication as framing [160, 161]. Politicization therefore can be understood as political framing by the usage of available choices that can create a political influence on readers.

Detecting politicization or political frames in articles is therefore a subjective issue because several factors can contribute towards framing. In order to understand framing in a meaningful way, several studies have measured different aspects of framing. Broadly, such factors include language construction by the media, and identifiable components of texts which people may perceive differently or with a different emphasis. Studies have identified frames using manual or human detection of frames. While manual approach is highly reliable due to the subjective nature of framing, it becomes a challenge when facing large sample size. Due to the nature of our dataset, we therefore rely on computer-assisted methods that are relevant to detect framing in large samples.

One strand of literature uses cluster analysis [162, 163] like topic models to capture framing. However the datasets used contain different topics, and so using this approach gives a fair advantage in separating different topics in the first attempt. Usually this is also assisted by manual annotations to improve the quality of detected frames. This approach of automated detection using LDA and similar models [164] for the data in this study can pose a challenge since the topic of this study is largely about climate change. (Indeed as expected, using the LDA model, we did not find substantial hints for distinguishing one topic from the other.) Although several sophisticated versions of clustering models exist in the literature, we do not use this approach because the basic LDA model does not seem to suit the data set we have and the objective we want to achieve (distinguishing potential political influences on readers).

Another strand of literature identifies factors that affect framing and attempts to measure such factors. This approach has the potential to separate different framing types even within a particular topic like climate change, as in this study. Framing does not directly alter the importance of the issue, unlike agenda setting [165]. However it can leave open the



door to subjectivity with potential to change individual beliefs about climate change and collective beliefs as people interact and discuss. Using a recent study, we find the following features to be the determinants of framing [166]: (i) sentence complexity representing overall organization and presentation of facts and opinions, (ii) presence of named entities, (iii) presence of phrases or sentences in quotes for emphasis, (iv) usage of images, and (v) catch phrases including headlines and tricky words. Most of these features can be identifiable using automated techniques. We emphasize the first two features for political framing in our analysis.

Politicization of articles in the Guardian can therefore arise from various choices made by authors in their articles in order to produce a desired way of presentation. In this study we investigate two choices of authors that can provide political contexts to (climate change) articles: (i) macro attribute determined by the categorization of articles in to different sections, and (ii) micro attributes composed of politically oriented entities. Both these factors are very intuitive in why they have the potential to impose different perceptions on readers. The macro attributes are more likely to encompass the tone and complexity of sentences readers might be expecting from articles in 'Politics' section, and especially the subtle differences experienced or new readers might feel when reading articles in the two categories: 'Politics', 'Others'. The micro attributes on the other hand are identifiable named entities. We believe these entities project political inclinations very well on to the readers if they are political terms. However such entities, even if are not precise political terms, might be susceptible to acquire political inclinations due to several contextual factors, one of them being their repeated presence/mentions in other articles which have high density of politically inclined terms. Quotes are recognizable and easily distinguishable aspects of text. Usually in newspapers, it contains statement made by different personalities/organizations, or points of emphasis. However we believe they carry a different notion than the presence of named entities and hence we prefer to keep this as a control variable in our analysis in next sections. In the Guardian, each article usually has a image or a few. We do not have precise information in our dataset about this, however we can

expect less variance in the number of images from one article to another and hence it may not be that useful to explain variances in discussion outcomes. Headlines usually contain named entities and we can expect these to appear again in the main text of the article (already captured by micro attributes). Apart from this aspect, headlines usually are written in a way to capture attention, and this is true for all articles' headlines. Therefore we don't see any reason to use headlines as a main instrument of politicization.

Based on the understanding so far, we adopt the following general formulation for the measure of politicization.

$$\text{Politicization} = \tilde{U} \left( \tilde{A}_1(\text{Macro features}), \tilde{A}_2(\text{Micro features}) \right) \quad (4.1)$$

We express politicization of an article as a function of the choices made by its author with respect to macro and micro attributes of the article and how readers perceive the political inclinations of such choices as they read the article. When readers decide whether to participate or not, it is based on their final perception of authors' choices. In (4.1),  $\tilde{U}$  is a function that shapes an article's interpretation by combining various macro and micro factors and interactions among them. At this stage we do not propose an exact formulation for  $\tilde{U}$ , however we would like our readers to know that any joint effects between macro and micro attributes that we test in later sections is to be understood as a part of the  $\tilde{U}$  function.

### 4.3.1 Macro attributes

The information on the section in which an article is published forms the macro attribute of politicization. Each article (about climate change) in Guardian belongs to exactly one section, the names of which are described in Table 3. We group the information about articles' sections into two relevant categories. As shown in Table 4, we have split all section names into two categories - 'Politics' and 'Others'.

What goes in to  $\tilde{A}_1$ ? At this stage, we have defined the macro attribute to be the category of an article - 'Politics' or 'Others'.  $\tilde{A}$  is supposed to capture all the complex phenomena that goes in to making one

category more politically inclined over the other category. Trying to provide a formulation for  $\tilde{A}_1$  would be beyond the scope of this study. What the formulation in (4.1) is instead saying is that the entire complex phenomena that can be attributed to  $\tilde{A}_1$  is a simple function of the category assigned to an article.

At this stage it is an untested assumption whether readers’ broader perception and hence follow-up responses in discussion depend on the article’s category or not. With the current formulation we are able to be in a position to only test in the later sections whether or not there is at least one (unknown or unquantifiable) factor that shapes political perception of article in a broader sense which can be characterized by the categorization of article.

It is important too understand that  $\tilde{A}_1$  does not look in to the content of articles. What it captures are those aspects of articles that are common across all articles in a given section/category. For instance, it can capture users’ expectations from an article in the ‘Politics’ section because seasoned readers might have been habituated with the use of language construction by authors when writing for this section.

**Table 4:** CATEGORIES FOR ARTICLES

Category	Number of Articles
Politics	3700
Others	48174

The section ‘Politics’ is assigned to ‘Politics’ category and all remaining sections are assigned to ‘Others’ category. Refer Table 3 for the frequencies of climate change articles under each section.

So to summarize, in this study, we do not make any assumptions about the formulation of  $\tilde{A}_1$ , mostly because we believe it is complex enough and beyond the scope of this research. That said, we also encourage future studies to investigate more on these aspects, especially using techniques from computational linguistics and psychology that can differentiate and quantify why different sections might be associated with

differing sense of perception, especially for articles related to same topic.

We shall later investigate empirically whether such unknown factors that shape interpretation of articles in ‘Politics’ section is different from that in other sections in the Guardian.

### 4.3.2 Micro attributes

As mentioned previously, we are looking at political inclinations of named entities in an article as the micro attributes for politicization. Inclusion of such entities are the choices of authors (and random to readers). There can be a lot of heterogeneity across how readers perceive different micro attributes. In this study, we base our analysis using minimal and important assumptions about temporal effect of perception which is facilitated by the memory of human minds and their ability to perceive contextual associations.

Political orientations of named entities form the gateway to politicization stemming from micro attributes of an article. We assume that political orientation of a particular named entity in a given article derives from two sources: (i) the direct political inclination captured by the entity and reflected to the readers which, on average, would hold without any past context about it mentioned in the Guardian, and (ii) the indirect political orientation of the entity perceived by readers assuming that entity has had a political context prior to being mentioned in the current article. An example for the first case can be the scenario in which “climate-change” is mentioned in an article of Guardian purely for the first time. An example for the second case can be a scenario where ‘climate-change’ is mentioned in its 101st publication (considering all articles) where the “climate-change” entity may have appeared multiple times in the past 100 publications along with named entities like “Republicans” and “Democrats”.<sup>1</sup> The indirect effect on perceived politicization is what we refer to as the *temporal effect of perception*.

The rationale for a temporal effect of perception is motivated from a psychological perspective which encompasses various complex factors

---

<sup>1</sup>The entities used for this example is solely for illustration and does not reflect any pre-conceived notions to influence the results of this study.

why a reader might perceive indirect political inclinations. While a non-political entity by itself may not carry any political orientation, it however may carry political notions due to its association or co-occurrence with political terms in the past. In a sense, the indirect political inclination may be understood as a kind of “network effect” on perception of readers when they come across named entities while reading an article. A simple illustration of this notion is shown in Figure 16.

To formalize a measure, we define the micro attribute of politicization of an article  $A_t$  published at time  $t$  as

$$PI^{\text{micro}}(A_t) = D \cdot \sum_{j \in E(A_{-t} \cap i)} w_j PI(j) + \sum_{i \in E(A_t)} PI(i), \quad (4.2)$$

where  $A_t$  represents an article published at time  $t$ ,

$D$  is a binary variable (0/1),

$E(A_{t'})$  is the set of unique named entities in an article  $A_{t'}$  published at a time  $t'$ ,

$PI(e)$  is the absolute political inclination of a named entity  $e$ ,

$A_{-t} = \{A_{t-1}, A_{t-2}, \dots\}$  is the set of all articles published prior to time  $t$ ,

$E(A_{-t} \cap i) = \{j \mid j \in E(A_{t'}), i \in E(A_{t'}) \ \forall \ A_{t'} \in A_{-t}\}$  is the set of all entities that co-occurred with entity  $i$  in articles published prior to time  $t$ , and

$w_j$  is a weight given by the cardinality of the previous co-occurrences of the  $(i, j)$  pair and normalized by the sum of all cardinalities of previous co-occurrences with the entity  $i$ .

Based on the value of  $D$ , we classify micro attribute into two types as mentioned below.

- *Direct type*:  $D = 0$  for this type, which represents that micro attribute of politicization of an article is derived only from the absolute political inclinations of all named entities in the article. Referring to  $\tilde{A}_2$  in (4.1), it suggests that the political perception derived from the named entities relies solely within the article and depends on each entity’s political inclination.

- *Indirect type*:  $D = 1$  for this type, which represents that micro attribute of politicization derives not only from the sum of absolute political inclinations of each entity, but also from the political inclination of each entity that is inheritable from that of previous entities with whom it co-occurred. Various formulations of inheritance are possible; for this study we choose a simple formulation that weights these inheritances by a normalized frequency of co-occurrences as shown in (4.2). Referring to  $\tilde{A}_2$  in (4.1), it suggests that readers' perceived politicization due to named entities in an article, as chosen by authors is a function of past contextual named entities. The contexts refer to all aspects of similarities/dependencies with past articles which readers perceive, and (speaking mathematically) is a function of the co-occurrences among entities, which we have quantified in (4.2).

**Table 5:** TYPES OF NAMED ENTITIES

Entity Type**	Meaning	Total number of appearances*
PERSON:	People, including fictional <sup>1</sup>	928407
NORP:	Nationalities or religious or political groups <sup>2</sup>	280839
ORG:	Companies, agencies, institutions, etc. <sup>3</sup>	912840
EVENT:	Named hurricanes, battles, wars, sports events, etc. <sup>4</sup>	18641
LAW:	Named documents made into laws <sup>5</sup>	9840

\*\* The text analysis tool Spacy is capable of recognizing these entities, and others, based on its trained model.

\* In the entire dataset, not just in the sample used for analysing this study

<sup>1</sup> For example, Obama, Cameron, Clinton, Trump, Abbott, Al Gore, etc.

<sup>2</sup> For example, Conservatives, Chinese, Democrats, Canadian, Syrian, Muslim, etc.

<sup>3</sup> For example, EU, Labour, Senate, White House, Google, World Bank, etc.

<sup>4</sup> For example, Hurricane Katrina, Hurricane Sandy, World War II, etc.

<sup>5</sup> For example, Kyoto Protocol, Climate Change Act, etc.

The only missing piece in (4.2) is a measure for the absolute political inclination of an entity. We find that there is a clear separation of entities' appearances across the two categories: 30,566 (unique) entities have appeared only in articles of the 'Politics' category, 432,332 named entities have appeared only in articles of 'Others' category, and 33,833 entities have appeared in articles of both categories. Using this information,

we define a simple measure of political inclination of an entity, irrespective of the time instances of its appearances in articles, as the number of times the entity has appeared in distinct articles of ‘Politics’ category in the entire dataset. Figure 16 illustrates the use of this measure towards computation of  $PI^{\text{micro}}(A_t)$ .

**Figure 16:** POLITICIZATION FROM POLITICAL INCLINATIONS OF ENTITIES



The figure shows entities from appearances across various categories in current article (the right box) and prior articles. The left box contains only those articles from the past which have co-occurred with entities of the current article at time  $t$ . The number inside brackets next to each entity is its absolute political inclination, i.e., the total number of times it has appeared in the Politics category in the entire dataset. Even if ent-1 never appeared in Politics category, it still contributes towards politicization (only in case of indirect type of micro attribute) because of its previous co-occurrences with p-ent-1 and ent-2. Next, ent-2 will contribute towards total politicization due to its own absolute inclination and also due to the inclinations from its previous associations with p-ent-1 and ent-3.

We make two choices for practical reasons in order to make the  $PI^{\text{micro}}(A_t)$  measure defined in (4.2) more meaningful and less compu-

tationally intensive. First, out of the various types of identifiable named entities that exist in the dataset, the ones we consider for the analysis are shown in Table 5. The reason is that these entities are more vulnerable to contain political inclinations. To explain with counter examples, named entity types like mountains, countries, products, etc. appears in many articles however including them in the measure for micro attribute of politicization is mostly irrelevant. For instance, we do not expect readers to perceive anything of political context as they come across various names of countries while reading an article. Second, to allow for computational efficiency while simultaneously maintaining a reasonable ground for audience’s ability to relate an article they are currently reading with previous articles, we restrict the set of past articles where we search for co-occurrences among entities to have been published within 60 days prior to the current article for which  $PI^{\text{micro}}(A_t)$  is being calculated.

## 4.4 Consumption perspective: Impact on collective discussion

In this section, we examine the influence of politicizing climate change articles on aggregate statistics of collective discussions in response to such articles. The identification of the impact of politicization in a complex discussion environment is largely possible due to the fact that named entities, storehouse of micro political inclinations, and category of a given article which decide the amount of political framing (or politicization) to which users are exposed to are exogenous to users because these are pre-determined and published prior to the discussion on the article. We also propose and show the statistical validity of two mechanisms that plausibly underlie the qualities of politicized discussions in response to politicized articles.

We have 22,171 observations after filtering the dataset for articles which are open for receiving comments and have actually received comments, as described in Section 4.2. (Of these, 1218 articles did not receive any comments.) The observations contain articles’ text, category infor-



mation, and their discussions for the climate change topic during 2006 to 2018.

Tables 18, 19, 20, & 21 show the consistency of the significance of coefficients estimated by regressing the aggregate outcomes of discussion on macro and micro attributes of politicization using different specifications. The models are estimated using the least squares method and heteroskedasticity robust standard errors. The macro and micro attributes are normalized<sup>2</sup> so that it can be ensured that the observed effects stem from relative changes within these attributes and not from their absolute magnitudes (which does not have an established conceptual meaning). The estimated coefficients of politicization attributes therefore should be interpreted only for their statistical significance and signs (positive or negative).

The estimates in Model 1 (across all the above tables) suggest that both macro and micro attributes significantly affect the discussion outcomes. These effects hold for both indirect and direct types of micro attributes, i.e., whether or not we might assume readers to be influenced by the political inclinations of previous articles.

Model 2 (across all the above tables) shows significant interaction effects between the macro and micro attributes. This suggests that, as expected, due to complex nature of framing, macro and micro attributes have significant joint roles in forming the perceptions of politicization that affect discussion outcomes.

Model 3 improves the reliability of estimates in Model 2 by controlling for the major potential confounders. As discussed previously in Section 4.3, three primary determinants of (political) framing include named entities, quotes, and sentence complexity. While named entities and sentence complexity are baked into the measures used for micro and macro attributes respectively, we control for quotes explicitly in two ways - the number of quotes used in articles, the length of all quotes measured by the number of words in all quotes. We also control for length of articles because this can affect both the perception of politicization and the content for discussion. On this note, we would like to mention that we have

---

<sup>2</sup> $x_I$ , the  $i$ th observation of a variable  $x$  becomes  $x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$  when  $x$  is normalized.

used only those variables as confounders which, by definition, have potential to affect both the discussion outcomes and the variables describing politicization.<sup>3</sup> The key takeaways from the above analysis done so far include the positive impacts of macro and micro attributes, and the significance of the interaction between macro and micro attributes.

Time (quarters of years) fixed effects are included in model estimates to eliminate bias from unobservables stemming from various characteristics of the time that are common across all articles within the time period. This makes the interpretation of the estimates less susceptible to time-specific events that happen within each period and can cause variations in authors' writeup of articles and in readers' attention from one time period to another.

Authors have objectives (e.g., agenda setting over a time period), use resources (e.g., named entities) and face constraints (e.g., article length) when they write articles. Therefore the choices of named entities across various articles might represent assignment of politicizations due to the authors. (In a treatment effects setting, this can be thought of assignment of treatments, the micro attributes, in clusters of articles grouped by authors.). This also allows to ensure robustness of estimates due to authors that are dropped out of the dataset used for regressions due several articles being not open for comments. Clustering standard errors by authors is statistically feasible because there are 6659 unique authors in the dataset. (There should be, ideally, large number of clusters.) For robustness, we also show standard errors clustered by authors for coefficient of macro attribute and interaction effect. In Tables 18, 19, 20, & 21, we see that the macro attribute may not be significant, especially with robust standard errors clustered by authors. The micro attribute (for both indirect and direct types) and the interaction effect remain significant.

So far, no discussion has been made about treatment effects exclusively. Estimates in Tables 18, 19, 20, & 21 do not allow for interpretation of counter-factual scenarios where one would be interested in understanding the impact of changing either the macro or the micro attribute.

---

<sup>3</sup>Do not control for post-treatment variables (Mostly harmless econometrics, Imai, Andrew Gelman pdf), peer influence paper

The reason these estimates do not allow for such causal interpretations are primarily because of the significance of the joint effects of macro and micro attributes which, according to framing theory and simple intuition, point towards the complexities in perception that can arise due to both positioning and mentioning articles with political orientations. Hence, the statistical significance of the impact on discussion outcomes arising solely due to a change in macro (or micro) attribute is unclear at this stage. However, expecting a positive impact (from both macro and micro attributes) seems to be plausible. We improve the causal interpretations of such impacts in the following sections, with a focus on justifying the conditional independence assumption.

Before proceeding, let us understand that there is no violation of the SUTVA, the stable unit treatment value assumption, in our analysis. For each article, we have the article's platform for publication on the web, there is an author who posts the content of the article, and there is a collective group of readers who join gradually to discuss the article. Each article, published at a given time, forms one unit (or one observation) in the statistical analysis. The author, the content, and the readers of an article are essentially features of the article about which we want to determine how one feature affects another. The treatment is the politicization of the article, either as a macro attribute or in the form of micro attribute. Hence treatment to an article is fully determined once its author has published it on the web, and strictly precedes any discussion on the article (thereby ruling out opportunities for reverse-causalities). Various aggregate characteristics summarizing the collective discussion of the article form potential outcome variables in our analysis.

SUTVA is least likely to be violated because of several reasons. *First*, politicization of an article, published by its author at a given time  $t$  might be related to the macro or micro attribute of politicization of articles published in the past and/or with other unpublished articles in the near future. There can be various reasons, for example, strategies by the author. However, this does not directly guarantee, and is least likely, that the treatment at  $t$  causes (or spills over to) politicization of future articles, or is influenced by that of past articles. To account for any similarities in

ideas by the author, we always ensure to use clustered standard errors as an additional layer of robustness check during estimations. This clustering is relevant if authors have been pre-determined at random as to how much each would politicize, and also if the audience has a habit in how they perceive the content of different authors.

*Second*, although an author’s choices about macro and micro attributes of an article at time  $t$  are exogenous to new viewers, nothing stops them from potentially spilling over their thoughtful responses on discussion sections of other articles, either of past articles since time  $t^-$  or of future articles upto time  $t^+$ . Plausible reasons for this can be that readers are usually exposed to a group of articles when they visit a site and they may be involved in driving comments across several articles. So what can be a good approximation for a duration for  $[t^-, t^+]$  during which *politicization of one article* might be, if any at all, driving *discussions on other articles* due to readers’ activities? By including quarterly fixed effects in regression models, we rule out event-specific and time-specific variations (including any common propaganda) so that any observed treatment effect of article at a time  $t$  can be ascribed to its own idiosyncratic politicization and not due to common shocks of politicization during the time period. Our estimates, to be seen later, suggest that the nature of inference of analysis does not change upon inclusion of time fixed effects, and hence further clustering is not necessary [167].

*Lastly*, it is noteworthy to mention that contextual associations with past articles to which readers might be responding to during discussions (as defined by temporal effect of perception) is conceptually different from violation of SUTVA, which requires that a large fraction of users should be writing their comments on engaging in discussion in other articles’ discussions after they have perceived politicization by reading a particular article.

#### 4.4.1 Impact of macro attributes

In this section we aim to detect if macro attributes play a role in affecting discussion outcomes by providing a causal analysis and interpretation.

An article may be written for a particular pre-determined section or the article may be assigned to a section after it has been written. In trying to investigate the impact of macro attributes, it is the later which forms the treatment. The correct interpretation is therefore a scenario where micro attributes are pre determined prior to such allocation of section. In this section, treatment refers to a *what-if* scenario of positioning an article from ‘Others’ category to the ‘Politics’ category, treated group refers to articles in the ‘Politics’ category, control group refers to articles in the ‘Others’ category, and treatment effect refers to the average treatment effect as in the Rubin Causal Model.

In order to improve model independence,<sup>4</sup> we use matching to preprocess the dataset. This will ensure estimated treatment effects are not sensitive to the model specifications used in the regressions and will minimize any potential violation of the conditional independence assumption to a large extent. This ensures that, on average, assignment of treatment to the control group is not accompanied by changes in confounders due to their correlations with the treatment. Exact matching can provide extremely unbiased estimates. In our attempt, we found a final dataset of only 72 observations, split equally between treatment and control groups. While 72 observations with high significance for treatment effect definitely provides reasons why estimates in previous section are not reliable (largely due to bias arising from correlations between macro and micro attributes which is now zero), it does not paint the full picture, especially about the authors. Unfortunately only 17 out of 6659 authors were retained in the reduced/matched dataset.<sup>5</sup>

We therefore choose coarsened exact matching (CEM) for reducing model dependence. CEM [168] is a monotonic imbalance bounding matching method that balances treated and control groups in a way that avoids the usual manual discovery of balanced dataset in which adjusting imbalance on one variable has no effect on the maximum imbalance of any other. Also, common empirical support, and robustness to mea-

---

<sup>4</sup>Model independence means that inferences made from any statistical estimation reveal underlying nature of the data, and that model specifications are not the cause of observed estimates.

<sup>5</sup>17 is the case for matching micro attribute of indirect type.

**Table 6: AVERAGE TREATMENT EFFECTS OF MACRO ATTRIBUTE**

	<i>Treatment: Macro attribute of politicization</i>	
	<i>Micro-attribute Type: INDIRECT</i>	<i>Micro-attribute Type: DIRECT</i>
<i>Effects (positive/negative/conditional) on various discussion characteristics</i>		
<b>Discussion Size</b> (Total Comments)	positive*	positive*
<b>Agreements/Feedbacks</b> (Total Likes)	positive*	positive*
<b>Users' Engagement</b> (Total Replies)	positive*	positive*
<b>Total Unique Users</b>	positive*	positive*

\* Estimates are significant with p-value less than 0.001.

Estimates are robust to heteroskedasticity, author clusters, time fixed effects, and potential confounders.  
For complete details, refer Tables 28, 29, 30, & 31.

surement error are guaranteed within the CEM methodology.

While a post-CEM dataset may reiterate the significance of the treatment effect as seen in previous section (and as seen with an exactly matched dataset, not shown here), it allows for more variation and hence the reliance of the estimates to be generalizable. Indeed, as shown in Table 22, there are 20,272 observations in the matched dataset; so we can expect much better unbiased estimates without having them to have high variance (as in the case of exact matching with only 72 observations). Also 6413 and 6323 authors out of 6659 are retained in the dataset when matched separately using micro attributes of indirect and direct types respectively; this helps enormously by making it feasible to cluster standard errors by authors during regression estimations. Overall, as shown in Table 22, a 98.6% mean improvement (considering all variables) in the mean difference across treated and control groups is obtained using the matching strategy. The estimates of treatment effect on various outcomes are briefly described in Table 6.

Tables 28, 29, 30, and 31 show the details of treatment effect of macro attribute of politicization on discussion size, social feedbacks, users' engagement, and unique users respectively. In Models 1, 2 (across all tables), we see that the estimates are consistent, and statistically signifi-

cant with positive impacts. The estimates are robust to heteroskedasticity, confounders, time fixed effects, and standard errors clustered by authors. It is sufficient to interpret the sign and significance since the numeric value of treatment effect estimate.

In Model 3 (for all collective outcomes), we see that there is heterogeneity in treatment effect. Hence the the effect of positioning an article in ‘Politics’ section on collective attention of readers would be higher if the article contains more mentions of politically oriented named entities.

How does the treatment channel? The macro attribute essentially captures organised patterns of writing, many aspects of readers’ perceptions, and probably a general influence of being political. It is difficult and beyond the scope of this investigation to propose how this might be happening. In any case, we can always assume it is for the same channel of perceptions for which political news are usually sensational on the media.

#### **4.4.2 Impact of micro attributes**

Since micro attributes (indirect and direct types) are continuous variables, we use a linear model to estimate their average treatment effects on discussion outcomes for each additional unit of change in the attribute of each type.

To improve model-independent estimates we use the Covariate Balancing Generalized Propensity Score (CBGPS) methodology [169], which is an extension of the covariate balancing propensity score (CBPS) methodology [170] to a continuous treatment. For a continuous treatment, CBGPS works by weighting observations to reduce the association between the treatment variable and covariates, so that a causal interpretation of the treatment effects is highly probable. CBGPS avoids the dichotomization of treatment variable, as one would do in propensity score matching, that can result in the loss of information [169]. We emphasize this concern since 22,171 observations may not be sufficient to test the robustness of dichotomization by varying the the threshold for the dichotomization. This would also risk splitting the already pre-matching unbalanced

macro attribute (1,165 and 21,006 observations in ‘Politics’ and ‘Others’ categories respectively) even further post-matching. More so, CBGPS comes with additional advantages which are not ignorable. Fong et al. [169] show that weighting observations derived from exactly identified CBGPS is more robust to misspecification compared to weights derived from recent alternative methods [171] and produces better balance of post-matched dataset.

Using the parametric estimation methodology for CBGPS (which performs equally well as the non-parametric version and is computationally faster), the correlations between treatment and macro attribute (and potential confounders) could be brought down to near zero values, as shown in Table 27. This definitely allows for better causal interpretations about the treatment effects on the discussion outcomes. As a part of the methodology, Box-Cox transformations of the treatments were performed before matching estimation in order to make their distributions closest to normal. (The optimal values for  $\lambda$  in the transformation were found to be -2 for both indirect and direct types.) The post-matched contains 22,171 observations with weights assigned to each observations from the CBGPS estimation. The estimates of impact of micro attributes is summarized in Table 7.

Direct treatment effects are not significant on total comments, social feedback and users’ engagement (Model 1 in Tables 28, 29, 30). These estimates are however significant under assumption of homoskedasticity (which is of less interest). The above treatment effects are however significant and also robust to clustered errors when macro attribute is included in regression (Model 2, 3 in Tables 28, 29, 30).

Direct impact of micro attribute, estimates in Model 1 of Table 31, is significant on the total users participating in the discussion. These estimates are significant irrespective of whether the macro attribute is controlled in the regression. Estimated coefficients are robust to heteroskedasticity, confounders, time fixed effects, and standard errors clustered by authors.

Identical estimates are observed for both indirect and direct types across all discussion outcomes. (This can be due to correlations between



**Table 7: AVERAGE TREATMENT EFFECTS OF MICRO ATTRIBUTES**

	<i>Treatment: Micro attributes of politicization</i>	
	<i>Micro-attribute Type: INDIRECT</i>	<i>Micro-attribute Type: DIRECT</i>
<i>Effects (positive/negative/conditional) on various discussion characteristics</i>		
Discussion Size (Total Comments)	conditional <sup>††</sup>	conditional <sup>††</sup>
Agreements/Feedbacks (Total Likes)	conditional <sup>††</sup>	conditional <sup>††</sup>
Users' Engagement (Total Replies)	conditional <sup>††</sup>	conditional <sup>††</sup>
<b>Total Unique Users</b>	positive <sup>†*</sup>	positive <sup>†*</sup>

\* Estimates are significant with p-value less than 0.001.

All estimates are robust to heteroskedasticity, and errors clustered by authors.

<sup>†</sup> Estimates are consistent only after accounting for potential confounders and time fixed effects.

<sup>††</sup> Estimates are positive, consistent and significant only after accounting for potential confounders and time fixed effects.

For complete details, refer Tables 28, 29, 30, & 31.

indirect and direct types of micro attributes.)

There are no robust confirmations, especially with clustered errors, for the heterogeneity of treatment effects across macro attribute (Model 3).

Due to entities-based approach to understand politicization in articles' main text, we are in a position to test two channels which we think might be the underlying mechanisms for collective discussion outcomes in response to micro attributes of politicization in the articles. (1) We saw above that the treatment effect on total users participating is significant. How many of them participate because they recognize political entities in comments? (2) How much of the total discussion outcomes for comments, social feedbacks, and users' engagement is driven by users who join with political influence? We investigate these two mechanisms in the following section.

### 4.4.3 Mechanisms of politicized discussions

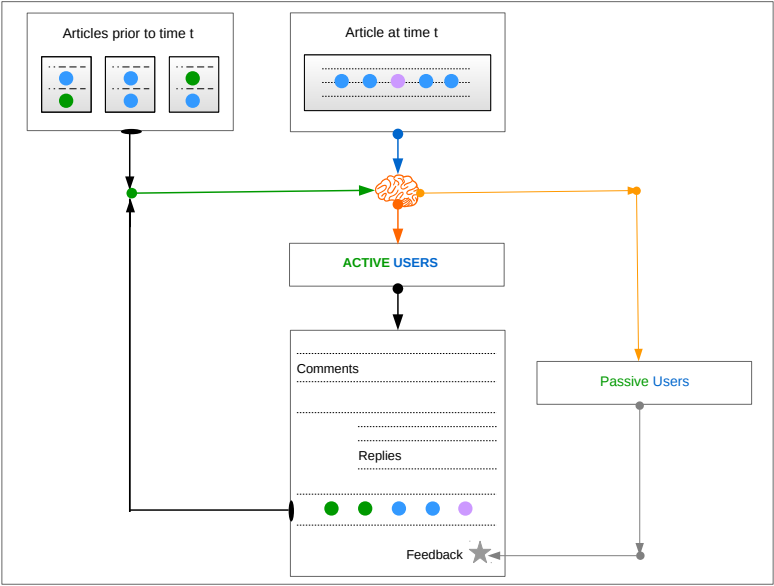
We begin with providing a description of the evolution of collective discussions, as illustrated in Figure 17, when users see an article.

Before that, let us take a quick note on why the estimates of the impact of micro attributes on various discussion outcomes had identical estimates. We find that indirect and direct types are actually positively correlated, which can be a potential reason. Also post CBGPS matching, with the requirement of Box-Cox transformation, we find that both these variables end up being highly correlated due to the nature of the transformation itself. The transformation requires both variables to be maximally correlated with the QQ plot, and at their optimal transformed versions, both were correlated with the QQ plot with a correlation coefficient of more than 0.9. While such a high value is indeed statistically desirable for the matching process, and hence having correct treatment effects for each type, it does not help to logically infer which of the two types might be a dominant factor that shapes readers' perception.

Let us understand the challenge in disentangling the source of effects (indirect vs. direct) with an example alongside the illustration in Figure 17. Suppose, as investigators, we see that the entity "blue" has appeared both in the current article and its discussion section, we might guess that the comments were influenced by the presence of such entities in the main text. When we perform treatment effects, as done in previous sections, we do it with the treatment being the political inclination of the "blue" entity for the direct type and the sum of political inclinations of the "blue" entity of current article along with the inclinations of "blue" and "green" entities derived from previous co-occurrences for the indirect type. (The "green" entity does not appear in the current article.) After the Box-Cox transformations performed as a part of matching methodology, it becomes difficult to distinguish from the identical estimates for indirect and direct types whether the effects arise from "blue" entities alone or from both "blue" and "green" entities. Logically, these two different sources of effects signify very different aspects of human behaviour, the later being the fact that there are effects coming from the memory because readers are able to form connections of the the current article with the past articles.

So, is there an actual effect from the memory? In order to answer this by distinguishing between the indirect and direct types of micro at-

**Figure 17: MECHANISMS FOR COLLECTIVE ATTENTION & DISCUSSION**



The dots represent named entities, carrying political inclinations. The green dots are of most interest because they appear in the discussion even if they are not present in the article at time  $t$ . Although absent in current article, the green dots have previously appeared in other articles along with blue dots, which are present in the current article. Some readers who are influenced by the politicization of the current article directly decide to comment about these by becoming active users (blue channel). A few other readers are reminded about green dots (either from their memory or by seeing the existing discussion - the green channel) and decide to become active users. At least 65% of readers turning into active users (and hence participating in discussion) due to micro attributes of politicization in the current article follow the green channel. The green-type users, who join the discussion due to political influence from outside contexts (green dots), drive at least a conservative estimate of 40% of the impact of micro attributes in current article on its total comments, agreements/feedback, and active users' replies to other's comments.

tributes, we extract the “green” entities from the combined set of “blue” and “green” entities that co-occurred in previous articles. We define political inclination of the discussions,  $PID$ , as the sum of political inclinations of entities that co-occurred (in previous articles) with entities of

current articles, are absent in the current article, and are present in the discussion of the current article. Thus, in reference to Figure 17, *PID* captures the political inclinations of only the “green” entities in the discussion section leaving out the political inclinations of the “blue” and “pink” entities in discussion.

Higher values of *PID* therefore not only suggest discussions becoming politicized, but also identify the source of such politically inclined entities to be strictly from previous articles that contain entities from current articles whose associations are perceivable by the readers. What are the chances that “green” entities appear in the discussion section? Well, they may appear by chance with some probability. This probability (of being random events) further decreases because these are not just some random entities from arbitrary articles in the past; instead these have precisely co-occurred with entities of current articles in previous articles, thereby suggesting indirect contextual associations between previous and current article via the common entities. (Such associations can manifest in discussions only if some readers actually perceive such contexts and discuss them in the comments.) More so, repeated appearances of “green” entities in articles over a dataset that spans over a decade definitely suggests that these occurrences are not random and definitely point towards readers’ ability to relate current article with the past. It is therefore better to investigate in the data whether statistically it has a role in shaping the collective discussion.

To summarize the measures for clarity, indirect micro attributes contain political inclinations of entities from both the current and past articles (“blue”, “pink”, “green”), direct micro attributes contain political inclinations of entities from current article (“blue”, “pink”), and *PID* contains political inclinations of only “green” entities. (Refer Figure 17.)

In previous section we saw that micro attributes of politicization significantly affect the number of unique users participating in the discussion. Users can have various reasons to actively join the discussion after being influenced by micro attributes of politicization (but before deciding whether to actively comment about it). In reference to Figure 17, the green channel is our interest of investigation.

- After a user is influenced by micro attributes, he may recollect “green” entities and decide to become an active user and comment about it, or he may see comments in the discussion section on “green” entities and decide to join the discussion. (The green channel does not distinguish between these two cases. Essentially it is an influence arising from “green” entities; either a user recollected from his own memory or saw it in discussion.)
- After a user is influenced by micro attributes, he may actively join the discussion to speak about “blue” or “pink” entities, or for any other reason from the set of unbounded number of possibilities. This is the blue channel.

Active users are therefore composed of two types depending on whether they decided to become active (i.e., to comment in discussion) via the green channel or the blue channel. What proportion of such users’ participation can be accounted to be catalysed by discussions becoming politicized (i.e., catalysed via the green channel)? In the language of causal mediation analysis, micro attributes form the treatment, politicized discussions as measured by *PID* forms the mediator, and total unique active users is the outcome variable. The average causal mediation effect (ACME) captures the treatment effect that mediates via discussions becoming politicized. The average direct effect (ADE) captures the treatment effect that mediates for all other reasons apart from discussions becoming politicized (i.e., via the blue channel discussed above).

Table 32 shows the estimates of ACME simulated using quasi-Bayesian Monte-Carlo method [172, 173], and White’s heteroskedasticity-consistent estimator. Estimates have been done for both indirect and direct types of micro attributes. We see that in Model 1, estimates of ACME are positive and highly significant for both types and suggest a value of 65% as the lower bound for the proportion of treatment effects mediated via discussions becoming politicized. Models 2 and 3 shows estimates of ACME conditional on the ‘Others’ category. The moderated mediation test is highly significant, suggesting that ACME estimates are substantially different across different categories (‘Politics’, ‘Others’). The ‘Others’ cate-

gory is more interesting since it shows that users tend to be influenced by politicized discussions (the green channel) even if the article is devoid of the potential political framing under ‘Politics’ category. The ACME estimates are quite consistent across model specifications of inclusion of confounders and interaction variables. Due to previously known joint effects between macro and micro attributes, this variable is controlled for. The confounders include all variables as used during estimating treatment effect of micro attribute (Refer Table 31).

All the above estimates undergo sensitivity tests [173] in order to assess the validity of the estimates under potential violation of the sequential ignorability assumption. As shown in Table 32, all estimations have 0.6 for the value of  $\rho$ , the sensitivity parameter which suggests the required correlation between error terms in mediator and outcome models before the causal mediation effect becomes zero. Indeed, 0.6 is a sufficiently high value and so the estimates are not sensitive to potential violation of the sequential ignorability assumption. (Refer Figures 24 and 25 for a visual understanding of the sensitivity analysis.) Also, the confidence intervals for the mediation effects in all models do not contain zero.

Therefore, based on above estimates, it seems that readers do indeed perceive “green” entities and a large fraction of users who are influenced by the micro attributes of politicization join the discussion after being further influenced by existing state of politicized discussions. While the magnitude of ACME estimates may not be directly interpretable, we at least know with high statistical accuracy that a substantial fraction of users, as large as 65%, join discussions being influenced by political inclinations of contexts (understood via entities) that were strictly not mentioned in the main text.

In Section 4.4.2, we saw that micro attributes do not directly affect the total comments, total social feedbacks, and users’ engagement. (Although these effects were significant under homoskedastic assumptions.) Since a user’s decision to participate in discussion follows after an article is published and before the user participates (actively or passively), it can be a potential reason why the treatment effects might not be di-

rectly significant. Hence users' participation seems to be an important mediating channel, and ideally we can expect the entire discussion to be mediated by this channel. (How else can a discussion emerge without users' participation?)

We therefore investigate how much of the discussion outcomes (comments, feedbacks, users' engagement) are mediated by users' participation, in particular by those users who are strictly influenced by political contexts not mentioned in the main article. We approximate the number of such politically influenced users by the predicted values of the regression of total (active) users' participation on *PID*. (Approximation using predictions from regressions is more sensible than directly assuming 65% of users as politically influenced users because previously discovered proportion of 65% speaks about distribution of the channelling of treatment effects. However the outcomes may be confounded by other variables, for e.g., the macro attribute, as we saw in the significance of moderated mediation test.) In the regression, we also control for macro attribute because this can affect both *PID* and total users. We also include squared terms for *PID* in order to control for potential non-linear effects.

Table 33 shows the ACME estimates for the effect of micro attributes on discussion outcomes mediated via participation of politically influenced users. We find that the most conservative estimate, without using matched dataset, have an average proportion (across various discussion outcomes) of at least 37% and 45% of mediation via such users for treatments of indirect and direct types respectively. Estimates from matched dataset in Models 2 & 3 suggest a highly significant mediation effect with non-significant direct treatment effects (ADE). The significance of these estimates suggest that the effect of micro attributes of politicization does indeed provide other politically oriented contexts to which users respond by participating actively, and such participation further drives the subsequent stages of evolution of discussion outcomes (comments, passive participation via feedbacks, engagement among users). Such contexts can be assumed to be coming from existing state of discussions for the treatment of direct type and from either past recollection or existing

discussion for the indirect type of micro attribute. (This deduction can be made using the definitions of direct and indirect types. For example, the direct type assumes readers to be influenced only by the political inclination of the current article and so any influence from outside contexts is more likely due to the existing discussions when such readers read an article. The types of micro attributes, indirect vs. direct, were not distinguishable in their impact on total users, as discussed previously using Table 26, but are now clearly distinguishable because of differing estimates of ACME, as seen in Table 33.) This also validates the assumption  $\tilde{A}_2$  in the sense that different mechanisms are in place when we assume different behaviours for readers with respect to their perception of political inclinations of entities they encounter in main article. (The difference in mechanisms arises from the probable source from which they recollect outside contexts to become politically influenced - own memory vs. collective imprints in discussions.)

## 4.5 Production perspective: Risk preference of authors

From authors' perspective, the total attention each article receives is random. Let us assume that if an author's article receives higher attention in terms of comments and participating users in its discussion, the author gets marginally higher satisfaction and that the marginal jump in satisfaction would be more during the initial moments of receiving readers' responses and their discussions. The author definitely did not pick all the politically oriented named entities in his article at random. The investigation in this section is to understand how sensitive is the author to choose micro attributes of politicization, holding other choices fixed, in a way that he is satisfied with the collective attention his article receives amid high uncertainty in collective participation. We do that by estimating the aggregate risk averseness of all authors.

We use an widely used model specification for estimating risk aversion in presence of risks to the factors driving satisfaction (collective attention in our case) [174], and use an alternative strategy that allows to



estimate the risk averseness coefficient without any assumption of the functional form of how satisfaction varies with collective attention. Also, there is no distributional assumption on the error term in model specification, as described below.

Suppose the satisfaction or utility  $U$  for an author from collective attention  $A$  to his article follows the assumptions as discussed above. We saw in previous sections that micro attributes of politicization positively impacts various outcomes of collection attention, however the extent of attention that would be received is uncertain. This uncertainty may vary depending on the level of micro politicization. (This is the reason why heteroskedasticity consistent estimates were reported during previous estimations of treatment effects using regression models.) Now, we assume that the collective attention  $A$  has the distribution

$$D(f(P, x), \tilde{g}(P, x))$$

where  $D$  is an unknown functional form, and for convenience can be characterized by its mean  $f(\cdot)$  and variance  $\tilde{g}(\cdot)$ , both of which depend on the level of micro attribute of politicization  $P$  chosen in the article and a vector of other attributes  $x$ . For estimation, we specify the following model

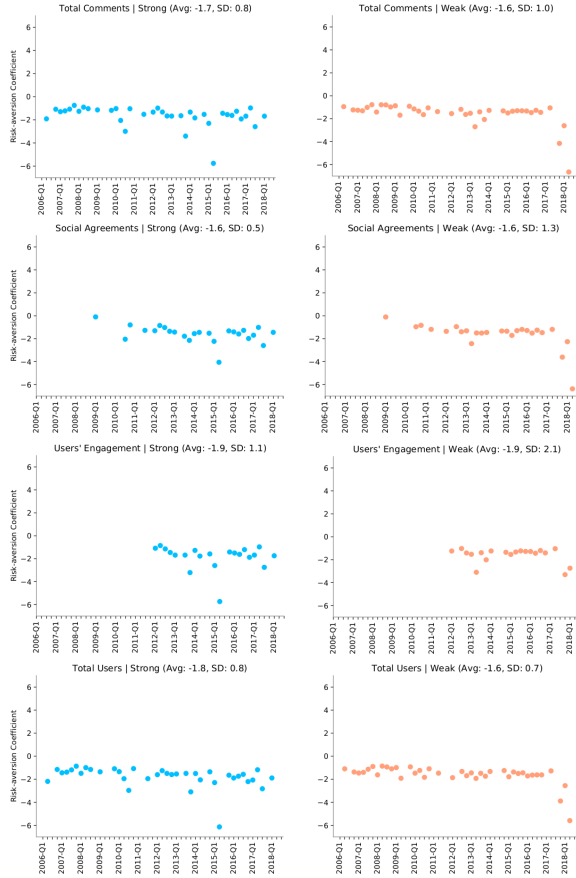
$$A = f(P, x) + g(P, x)\epsilon, \quad \mathbb{E}[\epsilon] = 0, \text{Var}(\epsilon) = 1,$$

where  $f$  is the mean attention function and  $g$  is the variance function which changes the uncertainty in attention for different levels of politicization  $P$ . The author would then maximize his expected utility from attention level  $A$ ,  $\mathbb{E}[U(A)]$ , by choosing the optimal level of  $P$ . The optimal level of  $P$  is determined by the first order condition of maximization

$$\frac{\delta f}{\delta P} + \frac{\mathbb{E}[U'\epsilon]}{\mathbb{E}[U']} \cdot \frac{\delta g}{\delta P} = 0, \quad (4.3)$$

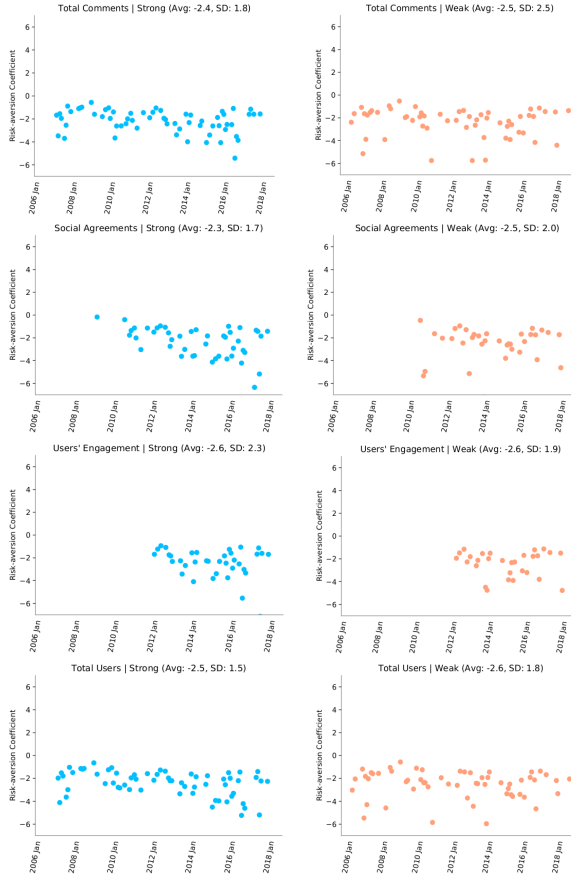
where  $\Theta = \frac{\mathbb{E}[U'\epsilon]}{\mathbb{E}[U']}$  is the risk preference function. Negative, zero, and positive values of  $\Theta$  would suggest risk-averse, risk-neutral, and risk-preferring attitudes on the author. (However, the magnitude of  $\Theta$  does not convey information about the intensity of risk-aversion. Only the

**Figure 18: RISK AVERSION COEFFICIENTS: QUARTERLY ESTIMATES**



The figure shows risk preference, estimated in quarterly windows, of authors in the Guardian for choosing politically oriented named entities in their articles to drive higher collective attention to their article on various outcomes. Missing points on the graph are due to insufficient data points or due to insufficient variations on values within the quarterly window to estimate the risk aversion. For all outcomes (across rows), we see that the estimates are below zero, suggesting that the authors are risk averse instead of risk-preferring to politicize climate change articles (for respective collective outcome, shown in each row). Across the columns, estimates are shown for indirect (or strong assumption) and direct (or weak assumption) types of micro attributes of politicization.

**Figure 19: RISK AVERSION COEFFICIENTS: MONTHLY ESTIMATES**



The figure shows risk preference, estimated in monthly windows, of authors in the Guardian for choosing politically oriented named entities in their articles in order to drive higher collective attention to their article on various outcomes. Missing points on the graph are due to insufficient data points or due to insufficient variations on values within the quarterly window to estimate the risk aversion. For all outcomes (across rows), we see that the estimates are below zero, suggesting that the authors are risk averse instead of risk-preferring to politicize climate change articles (for respective collective outcome, shown in each row). Across the columns, estimates are shown for indirect (or strong assumption) and direct (or weak assumption) types of micro attributes of politicization.

sign of  $\Theta$  is supposed to be inferred. The intensity can be inferred using further computations of absolute risk aversion, which is not necessary here.) Using this and equation 4.3, the risk preference function can be rewritten as

$$\Theta = -\frac{\delta f / \delta P}{\delta g / \delta P}. \quad (4.4)$$

We estimate  $\Theta$  in different time windows (quarterly and monthly) along the time. Within each time window, we construct uniformly separated bins for the micro politicization  $P$ , and estimate the mean  $f$  and variance  $g$  of attention  $A$  (various outcome characteristics) within each bin.  $\frac{\delta f}{\delta P}$  and  $\frac{\delta g}{\delta P}$  are estimated using slopes of the regressions of values of  $f$  from bins on the mean values  $P$  in bins. This piecewise approximation helps us capture the means and variances conditional on values of  $P$  and then the regression slopes estimate how they change with changes in  $P$ . The results are shown in Figures 18 and 19 for estimations of risk averseness using quarterly and monthly windows respectively. We see that authors are risk averse instead of risk loving to include more micro attributes in their articles and derive utility out of the total attention they receive on their articles. The conclusion is same for all variables that can hint at obtaining utility from online attention - total comments, total unique users, total social feedbacks, and users' engagement. So overall, the authors of the articles in Guardian have a risk averse behaviour.

## 4.6 Discussion

Politicization is an effect created due to authors' choices about articles' content and positioning, and readers' susceptibility to such choices whose effects are observed in how readers participate and engage in discussions. We saw that users are influenced by articles of climate change if those are politically framed, and it affects several key dimensions of collective discussions. We also saw that authors are not likely to politicize climate change articles for receiving more attention.

Comments and discussions are everywhere in digital spaces. Especially in the news media, it is important that public responds to the ob-

jective reality rather than how media presents the news. The collective perception on climate change (and any other issue in general) is what will drive participation and social actions leading to future policies. Hence we believe the results would be helpful for journalists, and the public in general.

*Limitations & Interpretations.* Since the study is based on observational data, users self select the articles they read and want to comment on. Although this would not affect the treatment (politicization of articles), there can be potential spillovers of treatments effects on other articles. We saw that the analysis is not sensitive to time-fixed effects (quarterly), and hence there is no need for further clustering. Hence we can assume that users mostly respond to politicization in by commenting on the articles where they discover it, and less likely on other articles.

Users may have different reasons to write or join discussion: articulation of personal opinions, reinforcement of or controversy with prior beliefs, promotion of agenda, emotional response, etc. In this study, we did not study the details of characteristics of comments which may help in revealing such motivations or other underlying mechanisms.

Our choice of named entities as the basic unit of framing is reasonable. This has been previously suggested in the literature and is also intuitive. Named entities form the least subjective factor among others like sentence complexity, existence of quotes, and others. Also named entities allow for understanding the links of a given article with articles published previously, not in a correlation manner but by being able to exactly point out the entities. Therefore, politicization of climate change due to micro attributes, as in this study, should be objectively understood as the effect on perception of climate change articles due to the presence of named entities of political orientation.

The choice of section name as a factor of politicization is also reasonable. Although finer details of underlying mechanisms are difficult to propose here, it definitely provides an overall sense of how the ‘Politics’ section distinctly differs from all other sections combined.

We believe that the results are less biased because we first showed the results using naive regressions, and then used better strategies for in-

vestigating a causal effect. Unobserved confounders which vary heterogeneously across different treatment levels are an obvious threat to the estimates, as in any non-experimental setting. However, in this study, sensitivity estimates in mediation analysis provides indirect ground for reliance on the estimates.

# Chapter 5

## Conclusion

### 5.1 Summary of Results

Human societies (i) carry a vast amount of information and make individual and group decisions, (ii) are highly connected for carrying out various activities, and (iii) are motivated for various complex reasons to do or not to do a particular activity. We may be less skilled, but after being influenced by our friends' skills, we may learn to become productive. We may be seeing tons of things on social media, but we may not be believing everything that we see. We may join to participate in a discussion not only because what the context is, but also how the context has been presented to us. It is a chaos, but the regularities within it are a wonder indeed. Understanding such regularities can help us build better societies, safeguard the environment, and improve economic activities. Digital platforms with social structures are a wonderful opportunity in today's age to attempt to infer such behaviour.

In this dissertation, we investigated some of such collective behaviour, each in a different digital society. We tried to understand effects on collective behaviour, and the mechanisms behind them. The main results of the contributions of this thesis are mentioned below.

- Peer influence of production popularity exists in the educational platform Scratch. Scratchers are influenced by their peers' produc-

tion popularity to create new projects which in turn increase their popularity. Such influence however is absent for Scratchers' consumption patterns. The causal effects of peer influence and mediation effects have been performed under quasi-experimental conditions with minimal behavioural assumptions in order to allow for generalizing the results to other platforms. Compared to the existing literature, it presents results on a widely adopted educational platforms and uses observational data over a long duration compared to live experiments, which may miss out on various complexities of users' behaviours that change over time.

- Credibility of fake information is shown to be an important factor which interacts with homophily in communication to create polarization of beliefs in a networked society. In the particular case of climate change sceptic messages in the social media Twitter, we found that such messages do not carry substantial credibility to polarize the beliefs of the society about the reality of climate change. Compared to the existing literature, it tries to emphasize the credibility of information, which has not been usually investigated in empirical studies on fake news and its propagation. Change of credibility can be a change of regime in how collective interactions and communications matter.
- Political framing, or politicization, of articles related to climate change in the news media Guardian is shown to influence the collective attention and participation of readers. Compared to existing literature, various underlying mechanisms have been shown with statistical validity, and the impacts have been estimated for articles from a single organization using its full historical data, thereby showing ways into detection of framing within a particular issue and within a given organization which using traditional methods, like improvements along the lines of topic models, may not sufficient.



## 5.2 Data is King & Context is Queen. They Dance Together!

I provide a short description of my perspective on experiments. Experiments and randomized controlled trials have been known to be the gold standards for causal inference. However, its applicability and generalizability need further judgements depending on the contexts.

If we are talking about a clinical trial, a particular drug tested in one part of the world has a high likelihood to work well across other societies and through time. If the trial is really experimental, this is exactly what is supposed to be inferred. If we are talking about a controlled or field trial to understand human behaviour, we can probably predict what is being written in the following lines. Yes, indeed. Individual behaviour and interactions among individuals leading to aggregate patterns of collective behaviour is a challenging thing to simulate in controlled environments. Field trials definitely project insights, however guaranteeing the results for different societies and in different times can be challenging unless the contexts and mechanisms are well understood. Unfortunately, human societies are highly evolved and complex, and so trying to infer something about us by experimenting on social behaviour among mice may not be useful. With monkeys, potentially yes!

It is all about the balance between bias and variance. Data collection and replication of results is one way, and being able to design methods for a new context and a new time is another way. Both of these should be encouraged in social science. Without a context and underlying insights on the collective behaviour, results may not be meaningful. To this end, observational data from digital societies (which pool in behaviour from several human societies) have advantages and disadvantages. Compared to the experiments, they can pose serious methodological challenges depending on the nature of data, have potential to answer more questions about complexities in collective behaviour, data collection can be expensive in certain situations, results can be inferred to be unbiased depending on how methods are employed, results may not be less biased than that in a controlled environment, and can have

greater scope for being generalized to unknown complex societies.

Nothing is without a fault, and nothing comes free (neither the data, nor the context). So depending on one's resources, one must choose between repeating randomized controlled trials in different contexts and conducting observational inferences in different contexts. Lastly, ethics matter. Whether it is experimental or observational, things like revelation of identities or conducting data extraction in secret ways are issues that bother humans. So we should be careful here. People may like to buy products advertised (and targeted) to them using partial identities revealed from collective behaviour, and they may consider to conduct clinical trials on mice to ensure their own safety, but they may not like to hear that their own government knows 'something' about them. I told you, human behaviour is complex!

## 5.3 Looking Ahead

There are several areas of opportunities to develop methods for analysing such behavioural data, and to analyse to understand collective behaviour that can help in improving social and economic policies in both physical and digital societies. Below is a list comprising a few potential areas for future research.

- *Digital privacy and ethics.* AI has provided several useful applications for the society. However, in the process of using large scale social data for such applications, issues like privacy breach and de-anonymization have raised concerns.
- *Roles of media.* Media in modern times, with massive participation of netizens, is playing several important roles in our societies like allocation of attention to particular topics, raising voice for justice in real time, making democracy more transparent, and others.
- *Epidemiology.* The coronavirus pandemic that started in 2020 has shown the importance of understanding disease spread in social networks, economies and their resilience, and proper communication during such disasters to be critical preparedness tools.

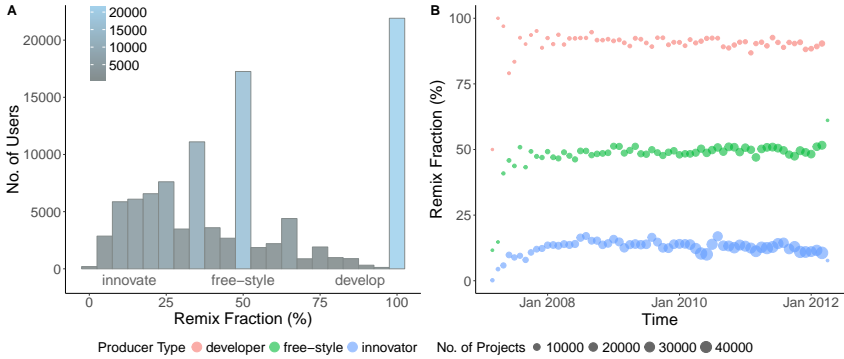
- *Future Technologies & Open Innovation.* Open source softwares in AI and big data age have shown that there is higher potential to innovate by recombining creativity, knowledge, experts, and communities. Transparent participation in digital innovation, creation of jobs, and effects on social and income equality need to be understood better. Health, agriculture, and education are promising areas to investigate further.

# Appendix A

## Peer Influence in Scratch

### A.1 Producer Types

Figure 20: TYPES OF PRODUCERS



(A) The distribution of remix fraction, i.e., percentage of remixed projects out of all projects created. There are 304,793 users who created at least one project. 202,018 users have zero remix fraction, and are not shown in the plot. The distribution is segmented into three types of producers: innovators, free-style producers, and developers. (B) The average remix fraction during each month for the three types of producers.

Here we provide a manual labelling of Scratchers as producers of

projects. Labels are based on their intensity to create remixed projects.

Remixing is a key feature of the Scratch community; it allows creation of projects based on existing ones. Fig. 20A shows the incentive of Scratchers to remix over the entire time duration; it is a distribution of the fraction of remixed projects among all projects created by a Scratcher. We use this distribution to understand if Scratchers lying in different parts of this distribution differ in certain behaviours. We create three definitions, based on the nature to remix projects: (i) developers: producers with remix fraction greater than 75 %. The value of 75 is chosen to increase the number of users in developers category; most of these users are in the top 10 percentile of the distribution. (ii) innovators: producers who mostly create new projects – producers with remix fraction less than 35 %, (iii) free-style producers: producers who do both. These users are the residuals of segmenting producers as developers and innovators. The value of 35 is chosen such that free-style producers, on aggregate, have 50% new projects and 50% remixed projects. Changes around the cut-off values of 75 and 35 does not affect the number of users in the interval very much. Innovators and developers have contributed to about 87% and 9% of new projects (non-remix projects) respectively.

The definitions are based on aggregate projects (and remixes) created in the entire duration. Users join the community over time, and they fall into one of these producer types (excluding non-producers) as defined by us by looking at the data of entire duration. To see if the labelling of producers based on production in the entire duration also holds in shorter intervals, we calculate the average remix fraction within each type in monthly windows as shown in Fig. 20B. It shows the average remix fraction of each producer type over time. We see that the group of producers categorized as developers (based on aggregate activity in 5 years) is of type developer in almost every month: in each month, the group produced projects that are mostly remixes. Developers did not behave as free-style type or innovator type in any month, except during the very early period. This suggests that the nature of producers is not volatile, and can be interpreted as a time-invariant behaviour. The distribution of the time spent on the platform by each of these types is almost

same, with an average of about 22 months and a standard deviation of 14 months.

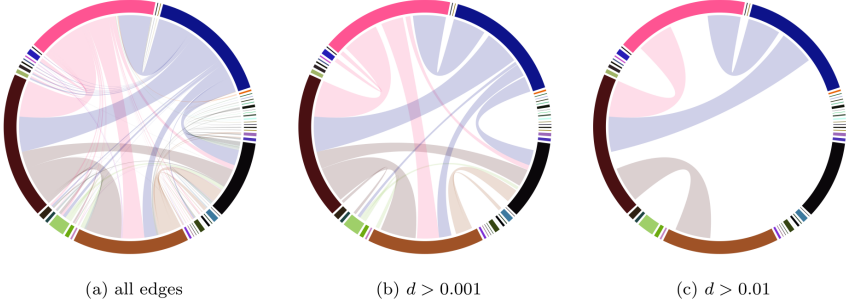
The volatility of free-style producers and developers in the early period can be explained by the fact that these producers need existing projects to remix. In the initial periods, since the online community was launched in 2007, there were less projects on board for these producer types to act by their nature. These types show a sharp deviation in favour of their nature, which is due to the availability of more projects in the platform due to passage of time.

Fig. 20A considers only non-zero remix fractions; users with exactly zero remix fraction are not shown. It forms a large fraction of all producers, however, we are not sure if such producers really did not remix at all. This is because we found some of these users to have produced large numbers of projects in comparison to others (outliers); such a situation might arise by copying projects [175]. Copying projects is legal, however it is unethical; copying is a situation in which a Scratcher modifies a non-substantial part of a project and then posts it as a new project and not as a remixed project (which references the original creator). Although copying can arise in other sections of the distribution in Fig. 20A as well, we did not find evidence of outlier cases for free-style producers and developers. In later analysis and discussions, we therefore study only these producer types: free-style producers, developers.

## A.2 Consumption Behaviour

Here we investigate if Scratchers, as consumers of projects, have preference to consume certain kind of projects. First we look for major consumption groups and next we investigate Scratchers' consumption specificity for such groups. Of the various forms in which users consume projects, only favorites and comments are non-anonymous records, i.e, in the available data, we can know the user who favorited or commented on a project. On the platform, favorites and comments are private and public information respectively.

**Figure 21: CONSUMPTION COMMUNITIES**



The communities in  $\mathcal{P}_{favorites}$  network, obtained by projecting users  $[u]$  on projects  $[p]$  nodes in the bipartite network where an edge  $u \rightarrow p$  represents  $u$  favorited  $p$ . Grids along the circumferences represent communities, and grid size is proportional to community size. Projects within each community are favorited together with high density. Projects of different communities are also favorited together but have low densities  $d$ , as shown by the edges between communities. Plots (a), (b), (c) have inter-community density  $d$  values higher than 0 (i.e., shows all links), 0.001, and 0.01 respectively.

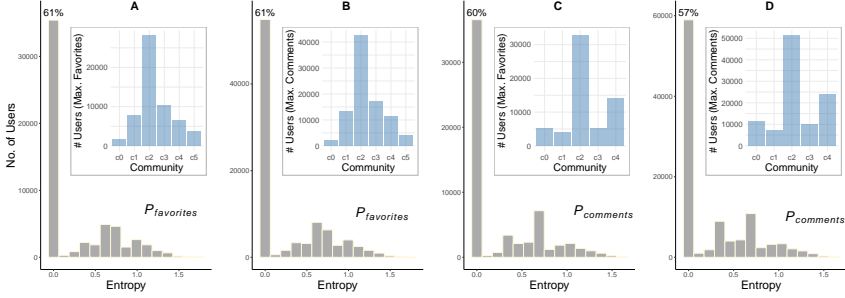
### A.2.1 Consumption Baskets

We consider a bipartite network of favoriting behaviour in which the nodes are users and projects, and edges are directed from each user to the projects which he has favorited. This is an aggregate network consisting of all favorites interactions in the 5 years. To see which projects are favorited together we obtain a bipartite projection on all projects; in the resulting network, an edge between two nodes (projects) has a weight equal to the number of Scratchers who favorited both the nodes. In the projected network,  $\mathcal{P}_{favorites}$ , there are 326,975 nodes and 162,611,378 edges with varying weights (ranging from 1 to 442). In the subset of  $\mathcal{P}_{favorites}$  with edge weights more than  $2^1$ , we found 145 communities in the network by implementing the Louvain algorithm [176]. (We performed communities detection using other algorithms as well, for example, we found 171 communities using fast greedy algorithm [177]. The

---

<sup>1</sup>This is done for computational simplicity and is inconsequential for analysis that follows.

**Figure 22: POLARIZED CONSUMPTION: EVIDENCE OF CONSUMPTION SPECIFICITY**



Main plots show distribution of entropy  $\mathcal{H}$  for consumption of types favoriting and commenting. In each case, there is a high fraction of  $\mathcal{H} = 0$ , meaning most users consumed projects exactly from a particular community. Insets show the distribution of users' maximal consumption group during 2007-12; for example, the inset in (A) shows that more than 20,000 users favorited projects from the  $c_2$  community in  $\mathcal{P}_{favorites}$  the maximum time. (A), (B) show distributions for  $\mathcal{P}_{favorites}$  and (C), (D) show that for  $\mathcal{P}_{comments}$  network.

main results that we discuss below is independent of the choice of algorithm.) 5 among the above 145 communities are of large sizes than others and the inter-community edge densities are low, as shown in Fig. 21. We perform a similar community detection on the bipartite network of commenting behaviour. In its projected network,  $\mathcal{P}_{comments}$ , there are 878,811 nodes and 1,097,722,712 edges, with edge weights ranging from 1 to 323. We found 4 large sized communities, using edge weights greater than 3.

We checked the tags of projects in each community to see if the projects across communities differ by particular topics. We found all communities have similar set of tags – game, simulation, animation, art, music, mario etc. – which are indeed very common tags on the Scratch platform. So the joint consumption of projects does not seem to be segregated by themes (as inferred by tags). We conjecture that the communities are formed by Scratchers' initial positioning in the friendship network upon joining the platform, and Scratchers in different segments of the network consume projects of similar themes.



## A.2.2 Consumption Specificity

We examine whether each Scratcher tends (intentionally or unintentionally) to consume projects only from specific communities found in previously.

We consider Scratchers who consumed (favorites, comments) projects from at least one of the 5 big communities, labelled  $c_1, \dots, c_5$ , found in  $\mathcal{P}_{favorites}$ . For each of these Scratchers, consider the distribution of consumption across  $c_0, c_1, \dots, c_5$  where  $c_0$  is the residual community of all projects not included in  $c_1, \dots, c_5$ . We measure a Scratcher's consumption polarization by an entropy-alike measure

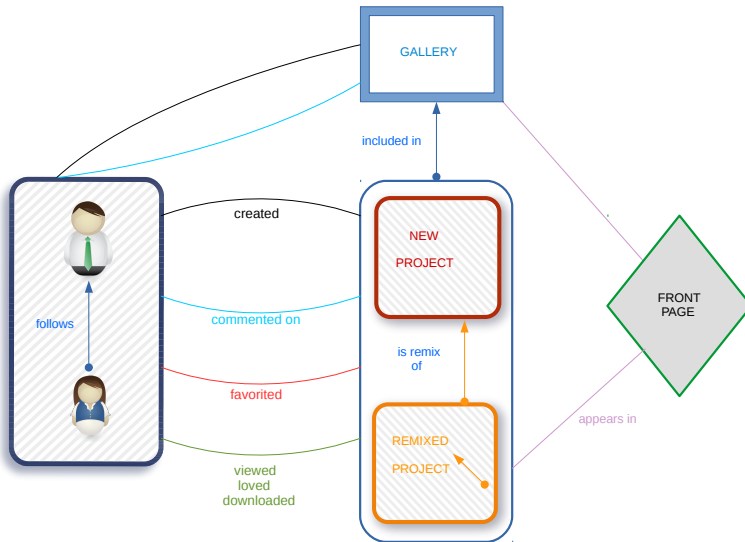
$$\mathcal{H} = -p_0 \log(f(p_0)) - \sum_{i=1}^n p_i \log(p_i), \quad n = 5;$$

$$f(p_0) = \begin{cases} 0.5 & \text{if } p_0 = 1, \\ p_0 & \text{if } p_0 \neq 1 \end{cases}$$

where  $p_i$ ,  $i = 0, \dots, 5$ , is the fraction of consumption from community  $c_i$  during the entire duration of five years. It is easy to verify that, with at least one positive  $p_i$ , the value of  $\mathcal{H}$  is 0 if and only if exactly one value of  $p_i$ ,  $i = 1, \dots, 5$  is 1. So if  $\mathcal{H} = 0$  for a Scratcher, he has consumed projects from exactly one of the 5 big communities. Fig. 22A and Fig. 22B show the distribution of  $\mathcal{H}$  values of all Scratchers for consumption of types favoriting and commenting respectively. About 60% of Scratchers favorite projects from only one community (Fig. 22A), and same is the case for commenting behaviour (Fig. 22B). We repeat this analysis considering the 4 big communities ( $n = 4$ ) found in  $\mathcal{P}_{comments}$ . As shown in Fig. 22C and Fig. 22D, we find evidence of polarization similar to the case for  $\mathcal{P}_{favorites}$ .

## A.3 Tables & Figures

Figure 23: SCRATCH PLATFORM



Users produce projects by creating and sharing on the platform. Projects are of two types – new and remixed. Users consume projects by commenting, favoriting, viewing, loving, and downloading them. Views, love-its, and downloads are anonymous. Users can follow each other, and form a friendship network. Projects can be included in galleries, created by shared users. Projects and galleries can be selected to appear on the front page.

**Table 8: VARIABLES DESCRIPTION**

Short Name	Description of Variable <sup>†</sup>
Love-its	Total love-its received on all projects created upto $t$
Views	Total views on all projects created upto $t$
Projects	Total projects created upto $t$
Galleries	Total galleries created upto $t$
Remixes	Total remixes (among all projects) created upto $t$
Age	Prior to $t$ , number of months in which ego interacted on the platform*
Active	True if ego interacted* > 1 month, prior to $t$
Favorited By	Total projects (of others) favorited by the ego (as a consumer) upto $t$
Comments (P) By	Total comments made by ego on (own and others') projects upto $t$
Comments (G) By	Total comments by ego on galleries upto $t$
Is Remixed	Of all projects created upto $t$ , total projects which have been remixed at least once (at any time)
Projects in Galleries	Total projects (of ego) appearances in various galleries created as of $t$
Front Page Projects	Total projects that appeared in front page upto $t$
Featured Galleries	Total galleries created by ego that were featured on the front page upto $t$
Studio Galleries	Total galleries of ego that appeared in studio design section upto $t$
Downloads	Total downloads (by others) of projects created by ego upto $t$
Favorites On	Total favorites (by others) of projects created by ego upto $t$
Comments On	Total comments received on projects created upto $t$

*continued on next page..*

**Table 9:** VARIABLES DESCRIPTION (..CONTINUED..)

Short Name	Description of Variable <sup>†</sup>
Featured Projects	Total projects that were featured on the front page upto $t$
Following	Total users the ego is following as of $t$
Followers	Total users who follow the ego as of $t$
Reciprocation	Total users who follow ego and are also followed by ego as of $t$
Peers Love–its	Total love–its received by all peers' projects upto $t$
Peers Views	Total views received by all peers' projects upto $t$
Peers Projects	Total projects created by all peers upto $t$
Peers Galleries	Total galleries created by all peers upto $t$
Peers Remixes	Total remixed projects created by all peers upto $t$
Peers Active	Total peers who have interacted* more than one month prior to $t$
Peers Fav By	Total favorites clicked by all peers upto $t$
Peers Is Remixed	Total projects of all peers upto $t$ which have been remixed at least once (at any time)
Peers Proj in Gall	Total projects (by all peers) appearances in various galleries upto $t$
Peers Fpage Proj	Total projects created by all peers which appeared in front page as of $t$
Peers Following	Total users all peers are following as of $t$
Peers Followers	Total users who follow any peer of the ego as of $t$

(<sup>†</sup>)  $t$  refers to a given point in time

(\*) Recorded forms of interactions only; does not include views, love-its, downloads because these interactions are anonymous. So if an ego stayed on the platform for only 1 month and downloaded many projects, his age evaluated at any future time is 0.

**Table 10: MODEL VARIABLES, CONFOUNDERS**

Objective: Dependent variable:	Determine Model Variables Change in Production Popularity <sup>†</sup>		Determine Confounders Treatment (1/0) <sup>††</sup>	
	[OLS Regression]		[Logistic Regression]	
	(1)	(2)	(3)	(4)
Peers Love-its ( $Tr^t$ )	0.341***	0.296***		
Love-its	0.077***	0.075***	0.011***	0.018***
Views	-0.003***	-0.003***	0.0001	-0.001***
Projects	0.021***	0.022***	-0.007***	0.007***
Galleries	-0.061***	-0.061***	0.144***	-0.051*
Remixes	-0.022***	-0.029***	0.020***	-0.001
Age	-0.024***	-0.010***	0.021***	-0.005*
Active	0.353***	0.128*	0.549***	0.150**
Favorited By	0.011***	0.007**	0.026***	0.008***
Comments (P) By	0.003***	0.003***	0.013***	0.003***
Comments (G) By	-0.0004**	-0.0002***	0.0001	-0.0002
Is Remixed	0.021***	0.020***	-0.006**	-0.009
Projects in Galleries	-0.0005	0.001		
Front Page Projects	0.286***	0.304***	-0.061***	-0.047
Following	0.005***	-0.041***	0.120***	-0.096***
Followers	-0.026***	-0.020***	-0.009***	-0.008*
Reciprocation	0.190***	0.237***	-0.695***	-0.186***
Peers Views		0.00002***		0.004***
Peers Projects		-0.001***		-0.010***
Peers Galleries		0.001		
Peers Remixes		-0.0002		
Peers Active		0.081***		0.140***
Peers Fav By		0.001***		0.003***
Peers Is Remixed		0.0001***		-0.007***
Peers Proj in Gall		-0.0003***		0.002***
Peers Fpage Proj		0.007***		-0.008
Peers Following		-0.0001***		0.00005
Peers Followers		-0.001***		-0.001***
Constant	-0.115*	-0.106	-1.165***	-4.381***
Observations	73,510	73,510	73,510	73,510
R <sup>2</sup>	0.296	0.307		
Adjusted R <sup>2</sup>	0.296	0.307		
Log Likelihood			-39,693.690	-7,285.876
Akaike Inf. Crit.			79,419.380	14,621.750
Residual Std. Error	7.365	7.304		
F Statistic	1,815.397***	1,164.492***		

Notes:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

This table shows regression results for  $t = \text{Dec } 2010, j = 1$ .

<sup>†</sup> Production Popularity: Love-its, <sup>††</sup> Treatment: Peers Love-its

**Table 11: BALANCE OF COVARIATES**

	Raw Sample	P. Score Match	(a) Exact Match (X)	(b) Exact Match (X,N)
<b>Panel 1</b>				
Love-its	17.65	17.65	0	0
Views	364.69	365.25	0.03	-3.23
Projects	13.42	13.42	0	-0.93
Galleries	0.8	0.8	0	0.01
Remixes	4.39	4.39	0	-0.24
Age	6.4	6.41	0	0.69
Active	0.22	0.22	0	-0.06
Favorited By	11.49	11.49	0	0.29
Comments (P) By	101.68	101.68	0	1.13
Comments (G) By	42.69	42.69	0.12	0.89
Is Remixed	3.69	3.7	0	-0.15
Projects in Galleries	12.11	12.12	-0.02	-0.11
Front Page Projects	0.27	0.27	0	-0.03
Following	17.04	17.09	0	-1.36
Followers	14.43	14.47	0	-1.09
Reciprocation	-0.08	-0.08	0	-0.13
Peers Views	107508.14	107507.79	16051.33	1757.93
Peers Projects	1454.85	1454.91	168.59	-6.92
Peers Galleries	68.34	68.41	7.1	0.12
Peers Remixes	453.48	453.48	52.26	-5.18
Peers Active	17.08	17.09	0.56	-1
Peers Fav By	1508.62	1508.61	160.76	18.2
Peers Is Remixed	1241.27	1241.28	163.41	0
Peers Proj in Gall	2108.5	2108.5	300.93	62.99
Peers Fpage Proj	68.73	68.73	9.59	0.93
Peers Following	2568.72	2569.78	296.73	0
Peers Followers	3326.23	3327.05	406.33	50.86
<b>Panel 2</b>				
Featured Galleries	0	0	0	0
Studio Galleries	0	0	0	0
Downloads	51.56	51.66	0	-0.55
Favorites On	12.77	12.78	0	-0.05
Comments On	103.69	103.73	0.05	-0.19
Featured Projects	0.02	0.02	0	0
<b>Panel 3</b>				
Treated Group	36697	36697	4502	2380
Control Group	36813	36697	11873	17570
Total Obs.	73510	73394	16375	19950

First panel contains variables used during regressions. Second panel contains variables that were not used for analysis due to one of these reasons: (i) correlated to love-its and carry similar information of a project's popularity, or (ii) multicollinearity detected during regression. Third panel contains count statistics.

Columns (in order) show the balance of covariates for  $t = \text{Dec 2010}$  before matching, after propensity score matching, after exact matching using  $X$ -type variables only, and after exact matching using  $X$ - and  $N$ -type variables. In (a), all variables are used for matching and in (b) only a subset of all variables is used (Exact matching on  $N$ -type variables is expensive.). Balance of an attribute is the difference of (weighted) means of the attribute between treated and control groups; weights are created in (a) and (b) when one user in treated group is matched to many users in control group.

**Table 12: PEER/NETWORK EFFECT**

<i>Dependent variable:</i> <i>Matched sample used:</i>	Change in Popularity (Love-its) of Projects			
	(a) Matched on X		(b) Matched on X, N	
	(Model 1)	(Model 2)	(Model 1)	(Model 2)
<b>Peers Love-its (<math>Tr^t</math>)</b>	0.017***	0.017***	0.037***	0.032***
Love-its		0.013		0.040***
Views		0.0003		-0.0003*
Projects		-0.001		-0.001
Galleries		0.001		-0.029***
Remixes		0.031***		0.014***
Age		-0.002***		-0.003***
Active		0.022***		0.035***
Favorited By		-0.001		-0.003***
Comments (P) By		0.003***		0.005***
Comments (G) By		-0.0001		0.0002
Is Remixed		-0.032		-0.002
Projects in Galleries		-0.011		0.004***
Front Page Projects		0.013		-0.012*
Following		0.002		0.002
Followers		-0.004**		-0.001
Reciprocation		0.016***		0.010
Peers Views	0	0	-0.00001***	-0.00000
Peers Projects	-0.00001	-0.00001	-0.0001	-0.0001
Peers Galleries	-0.0001	-0.0001	0.004***	0.004***
Peers Remixes	-0.00000	-0.00000	0.0002	0.0002
Peers Active	-0.0004	-0.001	0.002	0.0003
Peers Fav By	0	0	0.0002***	0.0002***
Peers Is Remixed	-0.00000	-0.00000	0.001	0.001
Peers Proj in Gall	-0.00001	-0.00001	0.0001***	0.0001**
Peers Fpage Proj	-0.00003	-0.00004	-0.003**	-0.002
Peers Following	-0.00000	-0.00000	-0.0002***	-0.0001
Peers Followers	-0.00001	-0.00000	-0.0001**	-0.0001***
Constant	0.004	-0.005	0.018***	-0.021***
Observations	16,375	16,375	19,950	19,950
R <sup>2</sup>	0.001	0.005	0.004	0.052
Adjusted R <sup>2</sup>	0.0003	0.004	0.004	0.051
Residual Std. Error	0.205	0.204	0.393	0.383
F Statistic	1.363	3.057***	7.255***	38.969***

*Notes:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

This table shows regression results for  $t = \text{Dec } 2010, j = 1$ .

# Appendix B

## Polarization in Twitter

### B.1 Cointegration Test

We have two time series processes during 2007-2017: polarization of beliefs and homophily in communication. Table 13 presents a naive look at the effect of homophily on polarization without adjusting the time series. In the models, homophily and lagged polarization are significant. Al-

**Table 13:** NAIVE LOOK AT RELATIONSHIPS IN TIME SERIES

	<i>Dependent variable: Polarization <math>P_t</math></i>		
	(1)	(2)	(3)
Homophily $H_t$	-0.202***	-0.153***	-0.162***
$H_{t-1}$		-0.007	-0.040
$H_{t-2}$			0.051
$P_{t-1}$		0.304***	0.262***
$H_{t-2}$			0.084
Constant	0.525***	0.381***	0.359***

*Significance:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

though polarization and homophily seem to follow a pattern, as shown in Figure 1 in the main text, inferring a relationship at this stage, as in Table 13, would be spurious due to autocorrelation in the curves.



We perform ADF test (using 2 lags) for checking presence of unit root; the null hypothesis of ADF test is that the corresponding time series is a  $I(1)$  process, i.e., it has unit root and hence is a non-stationary process. We find, as shown in Table 14, that polarization and homophily are integrated of order 1, while the residuals are stationary. We also perform Elliot, Rothenberg and Stock Unit Root test to arrive at the same conclusion. First difference of polarization, and first difference of homophily have ADF test statistics values of -13.55 and -15.1 respectively; this confirms that the differenced series are  $I(0)$ .

ADF test results can be sensitive to the number of lags used in the regression to obtain the residuals. Therefore, for a robustness check, we conduct Phillips-Ouliaris test using the  $\hat{P}_z$  statistic [178] which is not sensitive to the regression specification. The critical values for this test are available in Table IVa of Phillips and Ouliaris [178]. The critical values for  $\hat{P}_z$  statistic at 2.5% and 1% significance levels are 47.245 and 55.191 respectively. The  $\hat{P}_z$  statistic for the regression of polarization on homophily was found to be 53.9631; based on this evidence we reject the absence of cointegration at 2.5% level.

All the statistical tests conducted above suggest that polarization and homophily are  $I(1)$  processes while the residual of the regression of polarization on homophily is a  $I(0)$  process, thereby suggesting that polarization and homophily are cointegrated.

**Table 14:** ADF TEST STATISTIC ESTIMATES

	Type: None	Type: With Trend
Polarization	-0.1284	-3.2527*
Homophily	-0.0004	-2.6399
Residual	-3.9682***	-4.1087***
<i>Critical values (Trend):</i>	-3.99, -3.43, -3.13 (1%, 5%, 10%)	
<i>Critical values (None):</i>	-2.58, -1.95, -1.62 (1%, 5%, 10%)	
<i>Significance:</i>	* $p < 0.1$ ; ** $p < 0.05$ ; *** $p < 0.01$	

## B.2 Estimation using Vector Error Correction Model

With substantial evidence of the presence of cointegration, it becomes natural to model polarization and homophily jointly in a vector error correction (VEC) model [111, 112]. VEC models are a special class of models derived from vector autoregression (VAR) models with an additional term, called error correction term (ECT), to account for the cointegration and use first differenced variables instead of variables in levels as in a stationary VAR.

A lag length of  $p = 3$  was found to be optimum for a VAR model in levels based on the Akaike Information Criteria. Hence the lag length used for VECM model is 2 (one less than 3). The following equations (in vector form) are therefore specified as the VEC model for the investigation henceforth:

$$\Delta x_t = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \beta' x_{t-K} + \sum_{i=1}^2 \Gamma_i \Delta x_{t-i} + \epsilon_t, \quad (\text{B.1})$$

$$x_t = \begin{pmatrix} P_t \\ H_t \end{pmatrix}, \quad \Gamma_i = \begin{pmatrix} \gamma_{PP}^i & \gamma_{PH}^i \\ \gamma_{HP}^i & \gamma_{HH}^i \end{pmatrix}.$$

In equation (B.1), the term  $\beta' x_{t-K} = P_{t-K} + \tilde{\beta} H_{t-K}$  is the error correction term, which is the signature variable of a VEC model. The specification of VEC model in (B.1) are of two forms: long-run form and transitory form, which correspond to the cases where the values of  $K$  are 3 and 1 respectively.

Pfaff [179] provides an elegant demonstration of the steps to estimate VEC model. The first step in estimating the model (B.1) using the Johansen procedure [180, 181] consists of the Johansen trace test to determine the cointegrating rank  $r$  of the system. As shown in Table 15,  $r$  is found to be 1, i.e., there is one cointegrating relation, which is the maximum possible value for a system of two variables. (This also conforms to our previous conclusion about the presence of cointegration.) Next we

estimate the cointegrating vector  $\beta$ , and parameters in the VEC model by specifying cointegration rank of 1. The parameter estimates of model (B.1) are reported in Table 16.

**Table 15:** JOHANSEN COINTEGRATION TEST

Null Hypothesis	Trace statistic	Critical Values <sup>†</sup>
$r = 0$	18.77	(15.66, 17.95, 23.52)
$r \leq 1$	3.67	(6.50, 8.18, 11.65)

<sup>†</sup>Values in parantheses correspond to values of test statistic at 10%, 5%, and 1% levels of significance respectively.

**Table 16:** ESTIMATES OF VEC MODELS

	Transitory form	Long-run form
<i>Dependent variable: <math>\Delta P_t</math></i>		
$\mu_1$	0.15836*** (0.05280)	0.15836*** (0.05280)
$\alpha_1$ (ECT)	-0.28037*** (0.09372)	-0.28037*** (0.09372)
$\gamma_{PP}^1$	-0.32833*** (0.09981)	-0.60870*** (0.09280)
$\gamma_{PH}^1$	0.00846 (0.02809)	-0.06246* (0.03217)
$\gamma_{PP}^2$	-0.31197*** (0.08962)	-0.59234*** (0.10715)
$\gamma_{PH}^2$	-0.03162 (0.02725)	-0.10254*** (0.03734)
<i>Dependent variable: <math>\Delta H_t</math></i>		
$\mu_2$	0.31848** (0.12997)	0.31848** (0.12997)
$\alpha_2$ (ECT)	-0.55540** (0.23067)	-0.55540** (0.22841)
$\gamma_{HP}^1$	0.18988 (0.24566)	-0.36552 (0.22841)
$\gamma_{HH}^1$	-0.62945*** (0.06913)	-0.76994*** (0.07919)
$\gamma_{HP}^2$	0.21555 (0.22059)	-0.33985 (0.26372)
$\gamma_{HH}^2$	-0.14554** (0.06706)	-0.28603*** (0.09191)
<i>Significance:</i>	*p<0.1; **p<0.05; ***p<0.01	
<i>Note:</i>	Values in parentheses are standard errors of estimates.	

It is a recommended practice [182] to test whether the cointegrating vector satisfies either the restriction  $\beta' = (0 \ 1)$  or  $\beta' = (1 \ 0)$ ; these additional tests reduce the chances of spuriously concluding that near integrated variables are cointegrated. The likelihood ratio test statistic [179] for this hypothesis test has a  $\chi$  distribution with 1 degree of freedom in this case. The results of the test are given in Table 17 and, as can be seen, both restrictions  $\beta' = (0 \ 1)$  and  $\beta' = (1 \ 0)$  are rejected safely. Hence the conclusion about polarization and homophily being cointegrated remains valid and a long-run equilibrium relationship between them needs to be examined.

**Table 17:** HYPOTHESIS TESTS ON COINTEGRATING VECTOR  $\beta$

Restriction (Null Hypothesis)	Test statistic
$\beta' = (0 \ 1)$	11.42***
$\beta' = (1 \ 0)$	6.01***

*Significance:* \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

### B.2.1 Granger-Causality Tests

A stationary time series  $n_t$  is said to have a causal influence on another stationary time series  $m_t$  if the prediction of  $m_t$  can be improved by using lagged values of both  $m_t$  and  $n_t$  instead of using the lagged values of only  $m_t$ . The null hypothesis for Granger causality is that no explanatory power is added by jointly considering the lagged values of  $m_t$  and  $n_t$  as predictors. The null hypothesis that  $n_t$  does not Granger-cause  $m_t$  is rejected if coefficients of lagged values of  $n_t$  are significant, after having accounted for lagged values of  $m_t$ . For testing Granger-causality, the Wald test statistic follows the asymptotic chi-square distribution under the null. Since in our case, the time series of polarization and homophily are non-stationary (as seen before), the Wald test statistic does not follow the usual distribution.

To overcome the above constraint to use Wald statistic directly, we use the method suggested by Toda and Yamamoto [114]. According to

this method, additional lags (equal to the maximum integrating order in the system) are incorporated before conducting Wald test for Granger causality. Following this, we use a VAR model in levels (without differencing) using 4 lags. An additional lag (polarization and homophily are both integrated of order 1) is used on top of the optimal lag length of  $p = 3$  (determined previously using Akaike Information Criteria).

Testing homophily does not Granger cause polarization (null hypothesis) is performed using a Wald test where only the first 3 lags of homophily are restricted to zero. The Wald statistic follows a  $\chi^2$  distribution with 3 degrees of freedom under the null hypothesis. As shown in Table 1 in the main text, the null hypothesis is rejected (test statistic value = 12.3, p-value < 0.01). Testing polarization does not Granger cause homophily is performed by restricting the first 3 lags of polarization, and the test statistic is not sufficient to reject the hypothesis (test statistic value = 4.0, p-value > 0.1).

Hence the conclusion that emerges is that only homophily Granger-causes polarization with a negative effect, and the causality in the other direction is absent, i.e., polarization does not Granger cause homophily. Since the causality holds only in one direction, it is appropriate to interpret short term dynamics ( $\gamma_i$  in VEC model, (B.1)) only when evolution of polarization is the dependent variable. From Table 16 we see that the estimates of  $\gamma_{PH}^1$  and  $\gamma_{PH}^2$  are significant only in the long-run form of VEC model. Hence, homophily negatively affects polarization several months ahead.

### B.3 A Model of Polarization in Social Networks

Consider a social network with a uniform topology where each agent in the network (as a listener) has  $k$  speakers, neighbours from whom he receives information on a given topic (e.g., reality of climate change). An unobserved fundamental  $\theta \in \mathbb{R}$  describes the accuracy of the belief held by an agent regarding the topic: higher the  $\theta$  of an agent, the farther is the agent's belief away from the truth.<sup>1</sup> Suppose, at a time  $t$ , the agents in the social network are of two types -  $\theta_l, \theta_r$  - informed and misinformed, and the network has a homophily coefficient of  $\frac{h}{k}$  with respect to this property, i.e.,  $h$  speakers of an agent of type  $\theta_i$ ,  $i \in \{l, r\}$ , are of the same  $\theta_i$  type. Agents of types  $\theta_l$  and  $\theta_r$  represent two distinct sub-populations whose prior beliefs at time  $t$  are different: to formalize, we assume the former type have a prior belief about the fundamental  $\theta$  given by  $\theta_l \sim \mathcal{N}(\theta_0, \delta_l^{-1})$  while the later type have a prior  $\theta_r = \theta_l + \xi \sim \mathcal{N}(\theta_0 + \xi, \delta_r^{-1})$  which reflects the fact that they are misinformed. (Agents' prior beliefs are modelled as distributions [183], i.e., contain noise, to allow for heterogeneity of information sources based on which the beliefs are formed.) The population belief at  $t$  is composed of beliefs of the two sub-populations and is given by the mixture distribution

$$\theta_t^{pop} \sim f_l \cdot \theta_l + f_r \cdot \theta_r, \quad f_r = 1 - f_l,$$

where  $f_l$  and  $f_r$  are the fractions of the sub-populations of types  $\theta_l$  and  $\theta_r$  respectively. We define the belief of the population (social network) at time  $t$  to be *polarized* if the distribution  $\theta_t^{pop}$  has two modes. Mathematically,  $\theta_t^{pop}$  has either one mode or two modes [184].

Now we introduce communication among agents [185]. In the period between  $t$  and a future time  $t + 1$ , suppose each agent communicates his belief using a message (e.g., retweet a post to followers or mention to another user via a tweet), and all agents update their beliefs at  $t + 1$  after incorporating beliefs about the fundamental expressed in their speakers' messages. We assume that the credibility of communicated messages de-

---

<sup>1</sup>The choice of left or right side for denoting higher truth values does not affect the nature of results.

depends on the type of the speaker, i.e., in the social network, truthful and fake information propagate with different credibilities: listeners obtain an independent signal  $x_l | \theta_l \sim \mathcal{N}(\theta_l, \beta_l^{-1})$  from each speaker of type  $\theta_l$  and an independent signal  $x_r | \theta_r(\theta_l) \sim \mathcal{N}(\theta_r, \beta_r^{-1})$  from each speaker of type  $\theta_r$ . We assume that the messages from  $\theta_l$  type agents carry minimal credibility, i.e.,  $\beta_l$  is strictly positive. (In our context of the reality about climate change, this is a reasonable assumption.)

Lemma 1 characterizes the posterior beliefs of the two types of agents -  $\theta_l, \theta_r$  - at time  $t + 1$ .

**Lemma 1.** *The posterior belief of type  $\theta_l$  agents becomes*

$$\theta_l | \mathcal{I}_l \sim \mathcal{N} \left( \mu_l(\mathcal{I}_l), \frac{1}{\delta_l + h\beta_l + (k-h)\beta_r} \right) \text{ such that}$$

$$\mathbb{E}(\mu_l(\mathcal{I}_l)) = \frac{h\beta_l}{\delta_l + h\beta_l + (k-h)\beta_r} \theta_0 + \frac{(k-h)\beta_r}{\delta_l + h\beta_l + (k-h)\beta_r} (\theta_0 + \xi)$$

and that for type  $\theta_r$  agents becomes

$$\theta_r | \mathcal{I}_r \sim \mathcal{N} \left( \mu_r(\mathcal{I}_r), \frac{1}{\delta_r + h\beta_r + (k-h)\beta_l} \right) \text{ such that}$$

$$\mathbb{E}(\mu_r(\mathcal{I}_r)) = \frac{h\beta_r}{\delta_r + h\beta_r + (k-h)\beta_l} (\theta_0 + \xi) + \frac{(k-h)\beta_l}{\delta_r + h\beta_r + (k-h)\beta_l} \theta_0,$$

where  $\mathcal{I}_l$  and  $\mathcal{I}_r$  are the information sets comprising signals from neighbours for agents of types  $\theta_l$  and  $\theta_r$  respectively.

*Proof.* Below we sketch the proof for the distribution of  $\theta_l | \mathcal{I}_l$ , the proof for  $\theta_r | \mathcal{I}_r$  is similar. Consider a listener of type  $l$  who updates beliefs from two types of speakers,  $l$  and  $r$ . There are  $k$  neighbours of the speaker and among them  $h$  are of type  $l$  (homophily). We build the posterior for the listener in the following manner: he first updates information from speakers of type  $l$ , and then from speakers of type  $r$ . (This is same as obtaining the posterior using all signals at once.) With a prior  $\theta_l \sim \mathcal{N}(\theta_0, \delta_l^{-1})$ , the posterior obtained after updating with  $h$  signals  $x_l | \theta_l$  distributed  $\mathcal{N}(\theta_l, \beta_l^{-1})$  becomes

$$\mathcal{N} \left( \frac{1}{\delta_l + h\beta_l} \left( \theta_0 \delta_l + \beta_l \sum_{i=1}^h x_{l,i} \right), \frac{1}{\delta_l + h\beta_l} \right).$$



This becomes the prior for the next updates of  $k - h$  signals  $x_r | \theta_l$  distributed  $\mathcal{N}(\theta_l + \xi, \beta_r^{-1})$ . Hence the final posterior  $\theta_l | \mathcal{I}_l$  becomes

$$\mathcal{N}\left(\frac{1}{\delta_l + h\beta_l + (k-h)\beta_r} \left(\theta_0\delta_l + \beta_l \sum_{i=1}^h x_{l,i} + \beta_r \sum_{i=1}^{k-h} x_{r,i}\right), \frac{1}{\delta_l + h\beta_l + (k-h)\beta_r}\right).$$

In the case when  $\delta_l \rightarrow 0$ , this becomes

$$\mathcal{N}\left(\frac{1}{h\beta_l + (k-h)\beta_r} \left(\beta_l \sum_{i=1}^h x_{l,i} + \beta_r \sum_{i=1}^{k-h} x_{r,i}\right), \frac{1}{h\beta_l + (k-h)\beta_r}\right).$$

The expression for  $\mathbb{E}(\mu_l(\mathcal{I}_l))$  follows from the fact that  $\mathbb{E}(x_{l,i}) = \mathbb{E}(\mathbb{E}(x_{l,i} | \theta_l)) = \mathbb{E}(\theta_l) = \theta_0$  and similarly  $\mathbb{E}(x_{r,i}) = \theta_0 + \xi$ .  $\square$

**Lemma 2.** *The population is said to be polarized according to beliefs if the distribution of  $\theta_{t+1}^{pop}$  has two modes. Let*

$$\sigma = \left(\frac{\text{Var}(\theta_r | \mathcal{I}_r)}{\text{Var}(\theta_l | \mathcal{I}_l)}\right)^{1/2} = \left(\frac{\delta_r + h\beta_r + (k-h)\beta_l}{\delta_l + h\beta_l + (k-h)\beta_r}\right)^{1/2}, \text{ and}$$

$$\mu = (\mathbb{E}(\mu_r(\mathcal{I}_r)) - \mathbb{E}(\mu_l(\mathcal{I}_l))) \cdot \sqrt{\delta_l + h\beta_l + (k-h)\beta_r}.$$

Then  $f_l, \sigma$ , and  $\mu$  together determine whether the distribution of asymptotic population belief given by  $\theta_{t+1}^{pop}$  has one mode or two modes. In particular, if

$$\mu \geq \frac{3}{2}\sqrt{3} \min(1, \sigma), \quad (\text{B.2})$$

then the probability that the distribution of  $\theta_{t+1}^{pop}$  has two modes increases. If

$$\mu \leq 2 \min(1, \sigma), \quad (\text{B.3})$$

then the distribution of  $\theta_{t+1}^{pop}$  has exactly one mode.

*Proof.* This follows directly from Robertson and Fryer's Theorem on Modality [184]. The nature of the condition for polarization, as mentioned in (B.2), is probabilistic because additional conditions of large enough sub-population fractions  $f_l$  and  $f_r (= 1 - f_l)$  also need to be satisfied.  $\square$

**Proposition 3.** *In a random communication network with diffuse prior beliefs of both types of agents, polarization can not increase at time  $t + 1$ .*

*Proof.* (i) Using Lemma 1 and Lemma 2, when  $\delta_l \rightarrow 0$  and  $\delta_r \rightarrow 0$  (diffuse priors), we have

$$\mu = \xi \frac{\beta_r \beta_l (2kh - k^2)}{(h\beta_r + (k-h)\beta_l)(h\beta_l + (k-h)\beta_r)} \sqrt{h\beta_l + (k-h)\beta_r}. \quad (\text{B.4})$$

When the network is random, we can assume that on average each agent has equal number of speakers of the two types, i.e.,  $h = \frac{k}{2}$ . In this case,  $\mu$  is 0. Hence it follows from Lemma 2 that there is no polarization.  $\square$

**Proposition 4.** *If marginal increase in homophily at  $t$  does not positively affect the probability of polarization at  $t + 1$ , then the messages received from speakers of type  $\theta_r$  do not have a minimal level of credibility (i.e., a positive level of precision).*

*Proof.* For mathematical simplicity, we assume  $\theta_0 = 0$ . As we can see from the expressions of  $\mathbb{E}(\mu_l(\mathcal{I}_l))$  and  $\mathbb{E}(\mu_r(\mathcal{I}_r))$  in Lemma 1, the mean shift in the posterior belief from prior (e.g.,  $\mathbb{E}(\mu_l(\mathcal{I}_l)) - \theta_0$  for  $l$  type agents) depends only on  $\xi$ , i.e., to the extent of shift of mean belief, and does not depend on  $\theta$ , the mean of original belief.

**Case 1:** We first solve for the case when the priors are diffuse: we have  $\mu$  as given in (B.4). Using (B.2) and (B.3), we see that there is a higher probability for polarization to occur as  $\mu$  increases. If  $\mu$  does not increase when  $h$  increases (i.e., for fixed  $k$ , homophily increases), then there is no chance of polarization to grow since  $\mu$  has to be above a threshold for polarization ((B.2)) to happen. Hence, we use  $\mu$  as the parameter to gauge the probability of occurrence of polarization. Solving for the condition when homophily does not positively affect the probability of polarization, i.e.,  $\frac{\partial \mu}{\partial h} \leq 0$ , we have<sup>2</sup>

$$\begin{aligned} & \frac{2k\beta_l\beta_r\xi}{\sqrt{\beta_r(k-h) + \beta_l h(\beta_l(k-h) + \beta_r h)}} - \frac{\beta_l\beta_r\xi(\beta_l - \beta_r)(2hk - k^2)}{2(\beta_r(k-h) + \beta_l h)^{3/2}(\beta_l(k-h) + \beta_r h)} \\ & - \frac{\beta_l\beta_r\xi(\beta_r - \beta_l)(2hk - k^2)}{\sqrt{\beta_r(k-h) + \beta_l h(\beta_l(k-h) + \beta_r h)}^2} \leq 0 \end{aligned} \quad (\text{B.5})$$

For any positive distance away from the mean belief held by agents of informed type  $\xi > 0$ , for any positive number of speakers  $k > 0$ , for any non-trivial<sup>3</sup> level of homophily  $h \in (0, k)$ , and for any positive level of

<sup>2</sup>For easy replication, readers may use WolframAlpha online for solving.

<sup>3</sup>Trivial scenarios include cases where a listener has either all speakers of his own type or all speakers of opposite type. This is a purely mathematical restriction to avoid division by zero.

precision  $\beta_l > 0^4$  of informed messages, the only solution for (B.5) is  $\beta_r$  is 0.

**Case 2:** Now we solve for the general case when  $\delta_l$  and  $\delta_r$  are positive. In this case we have

$$\mu = \xi \frac{\beta_r h(\delta_l + \delta_r) - \delta_r \beta_r k + \beta_r \beta_l (2kh - k^2)}{(\delta_r + h\beta_r + (k - h)\beta_l) (\delta_l + h\beta_l + (k - h)\beta_r)} \sqrt{\delta_l + h\beta_l + (k - h)\beta_r},$$

and

$$\begin{aligned} \frac{\partial \mu}{\partial h} = & \frac{\xi(\delta_l + \delta_r)\beta_r + 2\xi\beta_l\beta_r k}{\sqrt{\delta_l + h\beta_l + (k - h)\beta_r} (\delta_r + (k - h)\beta_l + h\beta_r)} \\ & - \frac{(\beta_l - \beta_r)(\xi(\delta_l + \delta_r)h\beta_r - \xi\delta_r k\beta_r + \xi\beta_l\beta_r(2kh - k^2))}{2(\delta_l + h\beta_l + (k - h)\beta_r)^{3/2} (\delta_r + (k - h)\beta_l + h\beta_r)} \\ & - \frac{(\beta_r - \beta_l)(\xi(\delta_l + \delta_r)h\beta_r - \xi\delta_r k\beta_r + \xi\beta_l\beta_r(2kh - k^2))}{\sqrt{\delta_l + h\beta_l + (k - h)\beta_r} (\delta_r + (k - h)\beta_l + h\beta_r)^2}. \end{aligned} \quad (\text{B.6})$$

We can directly see that  $\frac{\partial \mu}{\partial h}$  is 0 when we plug the solution from Case 1, i.e.,  $\beta_r = 0$ . So even if the distribution already begins with some degree of separation of two peaks at  $t$  (since  $\delta_l$  and  $\delta_r$  are strictly positive in this case), the peaks do not further distance away at  $t + 1$ , i.e., irrespective of prior beliefs of agents, the credibility of fake information alone determines  $\frac{\partial \mu}{\partial h}$ . Now if  $\beta_r$  becomes positive, it is natural to expect that the beliefs of  $\theta_r$  type agents will be further reinforced, thereby strictly decreasing the spread of the right peak (i.e., increasing probability for polarization).

Indeed, looking for all solutions<sup>5</sup> of  $\frac{\partial \mu}{\partial h} \leq 0$ , under same restrictions as in previous case including  $\delta_l > 0$  and  $\delta_r > 0$ , we find that the only feasible solution is  $\beta_r = 0$ . Hence,  $\frac{\partial \mu}{\partial h}$  can not be strictly negative or zero when  $\beta_r$  is strictly positive.

So, in any case (Case 1, Case 2), a scenario where the increasing homophily does not increase the probability of polarization can only be explained when messages from type  $\theta_r$  speakers are not credible. (In other

<sup>4</sup>For the messages in favour of climate change, we assume that they carry a positive credibility.

<sup>5</sup>Verification is also possible by simulations. Using (B.6), it is easy to see that the sign of  $\frac{\partial \mu}{\partial h}$  depends on  $\beta_r - \beta_l$  which is basically the relative credibility of fake information with respect to true information. So with fixed values of  $\xi, \delta_l, \delta_r, k, \beta_l$ , we varied  $h$  from 0 to  $k$  and  $\beta_r$  from 0 to a very large value. We find that, irrespective of value of  $h$ , the value of  $\frac{\partial \mu}{\partial h}$  is strictly positive for  $\beta_r > 0$  and is 0 for  $\beta_r = 0$ . The code for this is available together with the code for empirical analysis, as mentioned in *Data availability* in the main text.

words, if fake information is at least slightly credible, the probability of polarization will always increase with increasing levels of homophily in communication.) ☐

## **Appendix C**

# **Politicization in Guardian**

### **C.1 Tables & Figures**

#### **C.1.1 Joint effects of macro and micro attributes**

**Table 18: EFFECT ON DISCUSSION SIZE**

Dependent variable: Total comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2	Model 3**	Model 1	Model 2	Model 3**
<i>Understanding perception of politicization via (i) effect of macro attribute, (ii) effect of micro attribute, (iii) interaction effect</i>						
Constant	18.043	146.960***	-50.435***	76.104***	181.598***	-54.303***
<b>Macro</b>	612.487***	-753.338***	-240.345*	612.434***	-1057.355***	-383.101***
(i)	(69.098)	(161.992)	(128.144)	(68.538)	(169.781)	(126.609)
<b>Micro</b>	2384.583***	1043.316***	247.249**	1693.002***	649.378***	294.560***
(ii)	(208.443)	(81.765)	(105.126)	(142.478)	(50.104)	(60.448)
<b>Micro × Macro</b>		4966.745***	3060.553***		4548.865***	2567.453***
(iii)		(719.268)	(581.453)		(573.077)	(440.710)
<i>Potential confounders, Robustness of estimates for politicization</i>						
Article Length			0.217***			0.217***
Total Quotes			-15.265***			-15.114***
#Words in Quotes			-0.2441***			-0.247***
Time FE	×	×	YES	×	×	YES
Author CSE (i)	0.000***	0.032**	0.397	0.000***	0.001***	0.165
Author CSE (ii)	0.000***	0.000***	0.072*	0.000***	0.000***	0.001***
Author CSE (iii)		0.001***	0.017**		0.000***	0.010
R <sup>2</sup>	0.112	0.150	0.341	0.105	0.154	0.343
Adjusted R <sup>2</sup>	0.112	0.150	<b>0.339</b>	0.105	0.154	<b>0.341</b>
F Statistic	71.66	107.9	120.5	73.65	108.2	120.8

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01  
Heteroscedasticity robust standard errors are reported in parantheses below the estimates.  
Micro attributes for politicization are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports only the p-value and significance of corresponding estimate when model is re-estimated using robust errors clustered by author names.  
Total observations: 22171 (21906 in models with Author CSE)  
(265 observations have missing author names)  
\*\* Final model used in main text for interpretation and further analysis

**Table 19: EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS**

Dependent variable: Total likes on all comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2	Model 3**	Model 1	Model 2	Model 3**
Understanding perception of politicization via (i) effect of macro attribute, (ii) effect of micro attribute, (iii) interaction effect						
Constant	109.417	689.413***	-257.735***	336.045***	812.694***	-291.539***
<b>Macro</b>	3221.590***	-2923.226***	-336.055	3080.951***	-4463.576***	-1116.422*
(i)	(340.925)	(732.968)	(583.578)	(336.851)	(779.225)	(598.620)
<b>Micro</b>	$1.1 \times 10^4$ ***	4961.886***	1672.092***	8213.798***	3498.443***	2169.658***
(ii)	(928.139)	(387.997)	(467.232)	(654.496)	(283.726)	(314.563)
<b>Micro × Macro</b>		4966.745***	2990.953***		$2.055 \times 10^4$ ***	$1.088 \times 10^4$ ***
(iii)		(719.268)	(593.688)		(2586.193)	(2021.607)
Potential confounders, Robustness of estimates for politicization						
Article Length			1.023***			1.007***
Total Quotes			-78.061***			-77.806***
#Words in Quotes			-1.112***			-1.098***
Time FE	×	×	YES	×	×	YES
Author CSE (i)	0.000***	0.075*	0.817	0.000***	0.003***	0.382
Author CSE (ii)	0.000***	0.000***	0.023**	0.000***	0.000***	0.000***
Author CSE (iii)		0.001***	0.021**		0.000***	0.012**
R <sup>2</sup>	0.105	0.136	0.348	0.103	0.143	0.352
Adjusted R <sup>2</sup>	0.105	0.136	<b>0.347</b>	0.103	0.143	<b>0.350</b>
F Statistic	81.11	113.0	79.19	85.20	108.4	80.58

Notes:

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Heteroscedasticity robust standard errors are reported in parantheses below the estimates.

Micro attributes for politicization are normalized.

Time fixed effects are quarterly fixed effects.

Author CSE reports only the p-value and significance of corresponding estimate when model is re-estimated using robust errors clustered by author names.

Total observations: 22171 (21906 in models with Author CSE)  
(265 observations have missing author names)

\*\* Final model used in main text for interpretation and further analysis

**Table 20: EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS**

<i>Dependent variable: Total replies to all comments</i>						
<i>Micro-attribute Type: INDIRECT</i>			<i>Micro-attribute Type: DIRECT</i>			
Model 1	Model 2	Model 3**	Model 1	Model 2	Model 3**	
<i>Understanding perception of politicization via</i>						
<i>(i) effect of macro attribute, (ii) effect of micro attribute, (iii) interaction effect</i>						
Constant	2.699	94.251***	-34.743***	41.481***	116.639***	-37.321***
<b>Macro</b>	446.750***	-523.196***	-163.109*	444.650***	-744.973***	-273.054***
(i)	(50.022)	(116.028)	(92.480)	(49.590)	(120.494)	(90.740)
<b>Micro</b>	1617.570***	665.063***	175.561**	1154.419***	410.901***	206.341***
(ii)	(148.357)	(58.887)	(70.616)	(101.237)	(36.593)	(41.458)
<b>Micro × Macro</b>		3527.157***	2210.192***		3240.790***	1873.800***
(iii)		(512.756)	(416.640)		(406.575)	(315.159)
<i>Potential confounders, Robustness of estimates for politicization</i>						
Article Length			0.149***			0.147***
Total Quotes			-11.746***			-11.623***
#Words in Quotes			-0.173***			-0.170***
Time FE	×	×	YES	×	×	YES
Author CSE (i)	0.000***	0.041**	0.433	0.000***	0.001***	0.165
Author CSE (ii)	0.000***	0.000***	0.049**	0.000***	0.000***	0.000***
Author CSE (iii)		0.001***	0.015**		0.000***	0.007**
R <sup>2</sup>	0.105	0.143	0.338	0.100	0.148	0.340
Adjusted R <sup>2</sup>	0.105	0.143	<b>0.336</b>	0.100	0.148	<b>0.339</b>
F Statistic	67.71	93.63	94.32	69.56	91.67	95.13

Notes:

\* p < 0.1; \*\* p < 0.05; \*\*\* p < 0.01

Heteroscedasticity robust standard errors are reported in parantheses below the estimates.

Micro attributes for politicization are normalized.

Time fixed effects are quarterly fixed effects.

Author CSE reports only the p-value and significance of corresponding estimate when model is re-estimated using robust errors clustered by author names.

Total observations: 22171 (21906 in models with Author CSE)

(265 observations have missing author names)

\*\* Final model used in main text for interpretation and further analysis



Table 21: EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION

Dependent variable: Total unique users						
Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT			
	Model 1	Model 2	Model 3**	Model 1	Model 2	Model 3**
Understanding perception of politicization via						
(i) effect of macro attribute, (ii) effect of micro attribute, (iii) interaction effect						
Constant	45.855	64.163***	-17.589***	56.354***	72.155***	-17.427***
Macro	124.080***	-69.884**	32.401	115.809***	-134.282***	-0.012
(i)	(14.423)	(29.237)	(23.803)	(14.210)	(30.208)	(23.434)
Micro	530.557***	340.080***	270.306***	400.611***	244.304***	222.784***
(ii)	(36.826)	(23.001)	(23.979)	(25.583)	(15.759)	(15.917)
Micro × Macro		705.341***	315.046***		681.300***	283.951***
(iii)		(120.120)	(97.008)		(95.648)	(73.948)
Potential confounders, Robustness of estimates for politicization						
Article Length			0.0352***			0.036***
Total Quotes			-2.415***			-2.363***
#Words in Quotes			-0.043***			-0.043***
Time FE	×	×	YES	×	×	YES
Author CSE (i)	0.000***	0.266	0.473	0.000***	0.018**	0.967
Author CSE (ii)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Author CSE (iii)		0.005***	0.140		0.000***	0.086*
R <sup>2</sup>	0.107	0.123	0.319	0.106	0.128	0.323
Adjusted R <sup>2</sup>	0.107	0.123	0.317	0.105	0.128	0.321
F Statistic	120.2	143.9	194.2	132.0	151.3	195.1

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01  
Heteroscedasticity robust standard errors are reported in parantheses below the estimates.  
Micro attributes for politicization are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports only the p-value and significance of corresponding estimate when model is re-estimated using robust errors clustered by author names.  
Total observations: 22171 (21906 in models with Author CSE)  
(265 observations have missing author names)  
\*\* Final model used in main text for interpretation and further analysis

### C.1.2 Impact of macro attributes

This section contains statistical tables showing the effect of macro attribute of politicization, the category assigned to articles ('Politics' or 'Others'), on collective attention/response.

**Table 22: BALANCE OF COVARIATES, SAMPLE SIZE, AUTHORS RETAINED**

	Before Matching	After Matching
<i>Micro-attribute Type: Strong</i>		
$\Delta$ Micro (normalized)	0.2451	0.0038
$\Delta$ Article Length	2617.0275	-5.8598
$\Delta$ Total Quotes	8.5677	0.7739
$\Delta$ #Words in Quotes	357.8501	11.5506
<i>Mean improvement of covariates in their difference across treated and control groups: 98.43%</i>		
Sample size of Treated group	1165	882
Sample size of Control group	21006	20272
Number of unique authors in dataset	6659	6413
<i>Micro-attribute Type: Weak</i>		
$\Delta$ Micro (normalized)	0.3452	0.005
$\Delta$ Article Length	2617.0275	-16.5518
$\Delta$ Total Quotes	8.5677	0.867
$\Delta$ #Words in Quotes	357.8501	21.2678
<i>Mean improvement of covariates in their difference across treated and control groups: 98.6%</i>		
Sample size of Treated group	1165	834
Sample size of Control group	21006	20025
Number of unique authors in dataset	6659	6323
<i>Matching method: Coarsened Exact Matching</i>		
<i>Treated group: Articles in 'Politics' category.</i>		
<i>Control group: Articles in 'Others' category.</i>		
$\Delta$ refers to mean difference of the covariate across treated and control groups.		

Table 23: EFFECT ON DISCUSSION SIZE

Dependent variable: Total comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
Treatment: Macro attributes						
Constant	344.8****	-36.6****	-32.84****	397.07****	-51.4****	-51.05****
Macro	477.8****	467.0****	-44.46	318.17****	333.62****	3.64
(i)	(71.9)	(58.9)	(110.63)	(93.97)	(56.8)	(122.2)
Micro		367.9**	279.75*		636.64****	597.74****
(ii)		(167.4)	(160.38)		(117.52)	(125.9)
Micro × Macro			2013.9****			998.21**
(iii)			(585.34)			(490.94)
Potential confounders, Robustness checks for treatment effect						
Article Length		0.12***	0.12***		0.12***	0.12***
Total Quotes		-6.57****	-6.71****		-7.56***	-7.58***
#Words in Quotes		-0.045	-0.039		-0.06	-0.06
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)	(241.6)**	(156.85)***	(125.83)	(205.49)	(117.87)***	(140.96)
Author CSE (ii)		(177.28)**	(181.32)		(129.6)****	(139.53)****
Author CSE (iii)			(856.6)**			(668.24)
R <sup>2</sup>	0.0137	0.335	0.340	0.005	0.426	0.428
Adjusted R <sup>2</sup>	0.0136	0.333	0.338	0.005	0.424	0.426
F Statistic	294.4	176.9	178.2	114.4	257.3	254.9

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

**Table 24: EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS**

Dependent variable: Total likes on all comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2 **	Model 3	Model 1	Model 2 **	Model 3
<i>Treatment: Macro attributes</i>						
Constant	1491.9****	-154.9****	-132.4****	1786.0****	-225.1****	-211.4****
<b>Macro</b>	2843.6****	2785.1****	-310.5	1970.8****	2018.4****	-185.26
(i)	(371.15)	(312.21)	(551.7)	(374.51)	(282.04)	(516.46)
Micro		2285.2****	1751.6**		3854.39****	3594.6****
(ii)		(786.34)	(764.65)		(497.12)	(520.16)
Micro × Macro			12189.6****			6666.2****
(iii)			(2906.9)			(2086.04)
<i>Potential confounders, Robustness checks for treatment effect</i>						
Article Length		0.33*	0.33*		0.21*	0.20*
Total Quotes		-18.32**	-19.13**		-17.81**	-17.97**
#Words in Quotes		-0.25	-0.21		-0.34	-0.33
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)	(1258.0)**	(905.84)****	(696.4)	(1045.42)*	(722.16)****	(775.69)
Author CSE (ii)		(894.39)**	(932.16)*		(587.4)****	(633.39)****
Author CSE (iii)			(4875.9)**			(3916.36)*
R <sup>2</sup>	0.017	0.287	0.294	0.008	0.3194	0.322
Adjusted R <sup>2</sup>	0.017	0.285	0.292	0.008	0.3175	0.320
F Statistic	364.6	141.7	144.0	163.6	162.7	162.2

Notes:

\* p < 0.1; \*\* p < 0.05; \*\*\* p < 0.01; \*\*\*\* p < 0.001

Heteroskedasticity robust standard errors are reported below the estimates.

Micro attributes are normalized.

Time fixed effects are quarterly fixed effects.

Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.

Total observations in each model: 22171 (265 missing author names),

Total observations when using errors clustered by authors: 21906.

\*\* Final model used in main text for interpretation and further analysis,

Model 3 only performs test for heterogeneity of the treatment effect.

**Table 25: EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS**

Dependent variable: Total replies to all comments						
Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT			
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
Treatment: Macro attributes						
Constant	216.2****	-23.09****	-20.3****	257.1****	-36.4****	-34.83****
Macro	350.89****	347.68****	-35.47	233.28****	250.76****	13.24
(i)	(52.08)	(42.87)	(77.84)	(70.51)	(42.17)	(91.47)
Micro		199.16*	133.11		402.99****	375.0****
(ii)		(107.51)	(106.37)		(86.56)	(93.14)
Micro × Macro			1508.78****			718.5*
(iii)			(413.91)			(367.10)
Potential confounders, Robustness checks for treatment effect						
Article Length		0.08***	0.08***		0.09***	0.09***
Total Quotes		-4.61****	-4.71****		-5.63****	-5.65****
#Words in Quotes		-0.02	-0.01		-0.04	-0.04
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)	(176.77)**	(116.2)***	(89.63)	(150.34)	(84.43)***	(103.31)
Author CSE (ii)		(111.91)*	(116.46)		(91.5)****	(99.0)****
Author CSE (iii)			(626.55)**			(486.26)
R <sup>2</sup>	0.015	0.351	0.357	0.005	0.441	0.442
Adjusted R <sup>2</sup>	0.015	0.349	0.355	0.005	0.439	0.440
F Statistic	313.8	190.3	192.1	109.4	272.9	270.2

Notes: \* p < 0.1; \*\* p < 0.05; \*\*\* p < 0.01; \*\*\*\* p < 0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

Table 26: EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION

Dependent variable: Total unique users						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
Treatment: Macro attributes						
Constant	139.51****	-14.12****	-13.83****	155.6****	-17.5****	-17.4****
Macro	107.55****	102.29****	62.53**	72.07***	73.8****	58.78**
(i)	(17.49)	(13.19)	(29.48)	(22.51)	(12.9)	(27.7)
Micro		230.31****	233.45****		274.8****	273.1****
(ii)		(54.29)	(52.48)		(32.8)	(34.6)
Micro × Macro			156.56			45.4
(iii)			(139.3)			(102.67)
Potential confounders, Robustness checks for treatment effect						
Article Length		0.025*	0.0257*		0.023**	0.023**
Total Quotes		-1.79***	-1.79***		-1.68**	-1.68**
#Words in Quotes		-0.02	-0.02		-0.02	-0.02
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)	(49.67)**	(26.56)****	(32.34)*	(44.6)	(21.66)****	(30.46)**
Author CSE (ii)		(62.14)****	(61.41)****		(41.31)****	(43.04)****
Author CSE (iii)			(171.71)			(123.39)
R <sup>2</sup>	0.007	0.329	0.329	0.003	0.398	0.3982
Adjusted R <sup>2</sup>	0.007	0.327	0.3272	0.003	0.396	0.3964
F Statistic	151.9	172.3	169.7	64.92	229.3	225.6

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

### C.1.3 Impact of micro attributes

This section contains statistical tables showing the effect of micro attributes of politicization on collective attention/response.

We begin with CBPS matching.. all authors are retained..

**Table 27:** BREAKING CORRELATIONS WITH CONFOUNDERS

	Before Matching	After Matching
<i>Treatment: Micro attribute (INDIRECT Type)</i>		
* Macro attribute	0.4517656	0.053593957
* Article Length	0.4772347	0.014060672
* Total Quotes	0.2745572	-0.003579448
* #Words in Quotes	0.2352519	0.017305180
<i>Treatment: Micro attribute (DIRECT Type)</i>		
* Macro attribute	0.4649682	0.05504194
* Article Length	0.3932794	-0.01314866
* Total Quotes	0.2452330	0.01315509
* #Words in Quotes	0.2048853	0.02138863
<i>Matching method: Covariate balancing propensity score (CBPS)</i>		
<i>Treatment: Micro attributes (Strong type, Weak type)</i>		
<i>Total observations: 22171</i>		
* refers to Pearson correlation between treatment and the covariate		

**Table 28: EFFECT ON DISCUSSION SIZE**

Dependent variable: Total comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
<i>Treatment: Micro attributes</i>						
Constant	330.19****	-60.3****	-59.45****	330.19****	-60.3****	-59.4****
Macro (i)		189.43**** (29.77)	78.43 (57.42)		189.43**** (29.77)	78.43 (57.42)
Micro (ii)	150.26 (389.28)	718.9**** (197.8)	700.78**** (202.21)	150.26 (389.28)	718.9**** (197.83)	700.79**** (202.21)
Micro × Macro (iii)			792.22** (358.81)			792.22** (358.81)
<i>Potential confounders, Robustness checks for treatment effect</i>						
Article Length		0.15****	0.15****		0.15****	0.15****
Total Quotes		-3.59****	-3.58****		-3.59****	-3.59****
#Words in Quotes		-0.14****	-0.14****		-0.14****	-0.14****
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)		(41.08)****	(70.99)		(41.08)****	(70.99)
Author CSE (ii)	(424.5)	(227.16)***	(232.3)***	(424.52)	(227.16)***	(232.3)***
Author CSE (iii)			(521.66)			(521.66)
R <sup>2</sup>	0.0002	0.505	0.505	0.0022	0.505	0.5052
Adjusted R <sup>2</sup>	0.0002	0.503	0.503	0.0018	0.503	0.5038
F Statistic	5.038	376.1	370.1	5.038	376.1	370.1

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* \* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.



**Table 29: EFFECT ON SOCIAL AGREEMENTS DURING DISCUSSIONS**

Dependent variable: Total likes on all comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2 <sup>**</sup>	Model 3	Model 1	Model 2 <sup>**</sup>	Model 3
Treatment: Micro attributes						
Constant	1595.57****	-357.9****	-355.6****	1595.57****	-357.9****	-355.6****
Macro (i)		1168.19**** (171.09)	857.61** (367.99)		1168.19**** (171.09)	857.61** (367.99)
Micro (ii)	1391.55 (1765.26)	4920.97**** (1041.55)	4870.3**** (1071.13)	1391.55 (1765.26)	4920.97**** (1041.55)	4870.3**** (1071.13)
Micro × Macro (iii)			2216.73 (2211.56)			2216.73 (2211.56)
Potential confounders, Robustness checks for treatment effect						
Article Length		0.73****	0.73****		0.73****	0.73****
Total Quotes		-19.19***	-19.19***		-19.19***	-19.15***
#Words in Quotes		-0.74****	-0.74****		-0.74****	-0.75****
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)		(232.9)****	(429.81)**		(232.9)****	(429.81)**
Author CSE (ii)	(2090.85)	(1314.6)****	(1348.48)****	(2090.85)	(1314.6)****	(1348.48)****
Author CSE (iii)			(3018.1)			(3018.1)
R <sup>2</sup>	0.0006	0.4647	0.4647	0.0006	0.4647	0.4647
Adjusted R <sup>2</sup>	0.00058	0.4632	0.4632	0.00058	0.4632	0.4632
F Statistic	14.07	319.8	314.6	14.07	319.8	314.6

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01; \*\*\*\*p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

**Table 30: EFFECT ON ENGAGEMENT IN DISCUSSION AMONG USERS**

Dependent variable: Total replies to all comments						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
<i>Treatment: Micro attributes</i>						
Constant	207.37****	-40.11****	-39.48****	207.37****	-40.11****	-39.48****
Macro (i)		141.12**** (20.95)	57.26 (38.33)		142.14**** (20.95)	57.26 (38.33)
Micro (ii)	90.85 (228.7)	508.75**** (115.24)	494.9**** (117.84)	90.85 (228.7)	508.75**** (115.24)	494.9**** (117.84)
Micro × Macro (iii)			605.66** (238.82)			605.66** (238.82)
<i>Potential confounders, Robustness checks for treatment effect</i>						
Article Length		0.09****	0.09****		0.09****	0.09****
Total Quotes		-2.49****	-2.49****		-2.49****	-2.49****
#Words in Quotes		-0.08****	-0.08****		-0.08****	-0.08****
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)		(27.98)****	(48.64)		(27.98)****	(48.64)
Author CSE (ii)	(253.72)	(130.79)****	(133.62)****	(253.72)	(130.79)****	(133.62)****
Author CSE (iii)			(376.3)			(376.3)
R <sup>2</sup>	0.0002	0.4897	0.4899	0.0002	0.4897	0.4899
Adjusted R <sup>2</sup>	0.00015	0.4883	0.4885	0.00015	0.4883	0.4885
F Statistic	4.381	353.7	348.1	4.381	353.7	348.1

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

Table 31: EFFECT ON TOTAL UNIQUE USERS IN DISCUSSION

Dependent variable: Total unique users						
	Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT		
	Model 1	Model 2**	Model 3	Model 1	Model 2**	Model 3
Treatment: Micro attributes						
Constant	88.58****	-25.39****	-25.37****	88.58****	-25.39****	-25.37****
Macro		54.54****	50.89****		54.54****	50.89****
(i)		(9.60)	(17.48)		(9.60)	(17.48)
Micro	287.95****	470.34****	469.7****	287.95****	470.34****	469.7****
(ii)	(64.5)	(39.11)	(39.9)	(64.5)	(39.11)	(39.9)
Micro × Macro			26.03			26.03
(iii)			(99.38)			(99.38)
Potential confounders, Robustness checks for treatment effect						
Article Length		0.02****	0.02****		0.02****	0.02****
Total Quotes		-0.14	-0.14		-0.14	-0.14
#Words in Quotes		-0.03****	-0.03****		-0.03****	-0.03****
Time FE	×	YES	YES	×	YES	YES
Author CSE (i)		(13.97)****	(21.18)**		(13.97)****	(21.18)**
Author CSE (ii)	(81.69)****	(62.37)****	(63.68)****	(81.69)****	(62.37)****	(63.68)****
Author CSE (iii)			(139.46)			(139.46)
R <sup>2</sup>	0.0145	0.349	0.3495	0.0145	0.349	0.3495
Adjusted R <sup>2</sup>	0.0145	0.347	0.3477	0.0145	0.347	0.3477
F Statistic	327.1	198.0	194.8	327.1	198.0	194.8

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01; \*\*\*\*p<0.001  
Heteroskedasticity robust standard errors are reported below the estimates.  
Micro attributes are normalized.  
Time fixed effects are quarterly fixed effects.  
Author CSE reports robust errors clustered by authors (in parenthesis) and significance for the corresponding estimates.  
Total observations in each model: 22171 (265 missing author names),  
Total observations when using errors clustered by authors: 21906.  
\*\* Final model used in main text for interpretation and further analysis,  
Model 3 only performs test for heterogeneity of the treatment effect.

#### **C.1.4 Mechanisms of politicized discussions**

**Table 32: POLITICIZED COLLECTIVE DISCUSSIONS: MECHANISM 1**

<i>Treatment: Politicization of article (Micro attribute)</i> <i>Mediation channel: Discussion becomes politicized by related but external contexts<sup>†</sup></i> <i>Moderated mediation via: Politicization of articles (Macro attribute)</i> <i>Outcome: Total unique users in discussion</i>						
	<i>Micro-attribute Type: INDIRECT</i>			<i>Micro-attribute Type: DIRECT</i>		
	Model 1**	Model 2 <sup>†</sup>	Model 3 <sup>†**</sup>	Model 1**	Model 2 <sup>†</sup>	Model 3 <sup>†**</sup>
Moderation Test		451.54****	514.84****		296.94****	371.01****
ATE	257.703****	250.74****	241.53****	223.82****	214.55****	242.532****
ADE	27.493	11.31	26.482	21.284	5.90	44.795
ACME	230.210**** [165.9, 289.7]	239.43**** [178.9, 304.9]	215.05**** [162.8, 268.2]	202.537**** [156.3, 251.5]	208.65**** [161.6, 253.0]	197.737**** [163.3, 232.8]
Sensitivity Test	(0.6, 0.209)	(0.6, 0.209)	(0.6, 0.209)	(0.6, 0.208)	(0.6, 0.208)	(0.6, 0.208)
Proportion	0.896**** [0.65, 1.38]	0.945**** [0.71, 1.52]	0.88**** [0.66, 1.39]	0.902**** [0.65, 1.43]	0.97**** [0.71, 1.53]	0.82**** [0.64, 1.11]
<i>Dataset and Variables Used for Model Specification</i>						
Matched Dataset	YES	YES	YES	YES	YES	YES
Micro × Macro	YES	YES	YES	YES	YES	YES
Mediator × Macro	×	YES	YES	×	YES	YES
Confounders	×	×	YES	×	×	YES

*Notes:* \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001. Square brackets report confidence intervals. Total observations in each model: 22171. Models are estimated using quasi-Bayesian approximation and heteroscedasticity robust standard errors. Sensitivity Test reports values of  $\rho$  and  $\bar{R}_M^2$ ,  $\bar{R}_Y^2$  (in order) within parantheses.  $\bar{R}_M^2$ ,  $\bar{R}_Y^2 = \rho^2$ . Matched dataset is the one obtained during treatment effect estimation. See Table 27. Macro variable is a binary variable for article's category: 1 for 'Politics' 0 for 'Others'. Confounders used in models are article length, total quotes, and number of words in quotes. Moderation Test reports difference in ACME values for estimates conditional on category (Macro attribute), i.e., ACME('Others') - ACME('Politics'), and whether the difference is statistically significant. <sup>†</sup> Refers to discussion of an article being populated with politically inclined entities which (i) have appeared in previous articles that co-occurred with entities of the current article but (ii) have never appeared in the main text of the current article. <sup>†</sup> Model 2 and Model 3 show estimates conditional on 'Others' category. \*\* Final models used in main text for interpretation and further analysis

**Table 33: POLITICIZED COLLECTIVE DISCUSSIONS: MECHANISM 2**

Treatment: Politicization of article (Micro attribute)						
Mediation channel: <b>Users influenced by related contexts but external to current article</b> †see next page						
Outcomes: Comments, Feedbacks, Users' Engagement						
Micro-attribute Type: INDIRECT			Micro-attribute Type: DIRECT			
	Model 1	Model 2	Model 3**	Model 1	Model 2	Model 3**
Outcome: Discussion size (as in Table 18)						
ATE	1304.846****	206.11	205.88	888.818	235.1	252.05
ADE	822.26****	-471.41	-471.43	491.398****	-393.3	-375.15
ACME	482.586****	677.52****	677.31****	397.42****	628.4****	627.2****
	[424.4, 542.3]	[490.3, 876.3]	[480.8, 900.5]	[354.0, 443.2]	[465.8, 803.6]	[457.6, 811.0]
Sens. Test	(0.4, 0.1133)	(0.6, 0.1535)	[0.6, 0.1535]	(0.4, 0.1124)	(0.6, 0.1523)	(0.6, 0.1523)
Proportion	0.371****	1.18	1.22	0.45****	1.4	1.39
Outcome: Social agreements during discussion (as in Table 19)						
ATE	6140.5****	1252.8	1256.8	4577.1****	1881.4	1888.18
ADE	3479.02****	-2553.7	-2546.9	2408.7****	-1602.7	-1609.8
ACME	2661.5****	3806.6****	3803.8****	2168.3	3484.1****	3498.1****
	[2348.5, 2974.0]	[2869.2, 4914.1]	[2763.6, 4966.2]	[1918.4, 2427.8]	[2521.3, 4494.0]	[2538.7, 4479.9]
Sens. Test	(0.4, 0.1125)	(0.4, 0.0743)	(0.4, 0.0743)	(0.4, 0.1112)	(0.4, 0.0738)	(0.4, 0.0738)
Proportion	0.434****	1.45	1.51	0.475****	1.28	1.37
Outcome: Engagement in discussion among users (as in Table 20)						
ATE	849.2****	98.23	100.39	579.9****	139.09	146.16
ADE	515.37****	-356.5	-354.25	305.9****	-280.54	-275.37
ACME	333.8****	454.72****	454.63****	274.1****	419.63****	421.5****
	[289.9, 376.2]	[340.3, 582.4]	[340.0, 583.1]	[243.5, 307.5]	[309.0, 530.7]	[321.2, 538.6]
Sens. Test	(0.4, 0.1152)	(0.6, 0.1555)	(0.6, 0.1555)	(0.4, 0.1141)	(0.6, 0.1543)	(0.6, 0.1543)
Proportion	0.393****	1.59	1.71	0.472****	1.56	1.67
Dataset and Variables Used for Model Specification						
Matched Data	×	YES	YES	×	YES	YES
Micro × Macro	YES	YES	YES	YES	YES	YES
<sup>1</sup> Med × Macro	YES	×	YES	YES	×	YES
Confounders	×	YES	YES	×	YES	YES

Notes: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01; \*\*\*\* p<0.001. Square brackets report confidence intervals. Total observations in each model: 22171. Models are estimated using quasi-Bayesian method and heteroscedasticity robust standard errors.

Sens. Test reports sensitivity test values of  $\rho$  and  $\bar{R}_M^2$ ,  $\bar{R}_Y^2$  (in order) within parantheses.  $\bar{R}_M^2$ ,  $\bar{R}_Y^2$  =  $\rho^2$ .

<sup>1</sup> Med refers to the mediator variable.

Matched dataset is the one obtained during treatment effect estimation. See Table 27.

Macro variable is a binary variable for article's category: 1 for 'Politics' 0 for 'Others'.

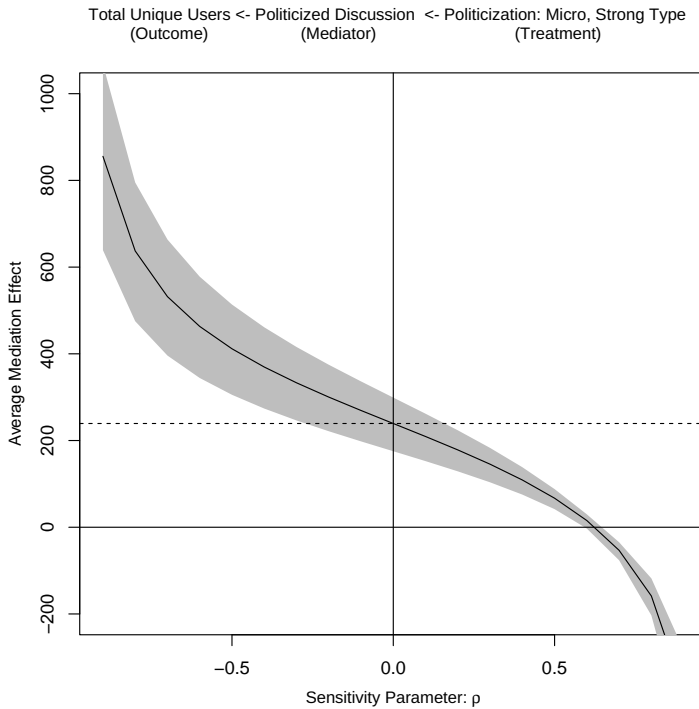
Confounders used in models are article length, total quotes, and number of words in quotes.

\*\* Final model used in main text for interpretation and further analysis

*Additional Note for Table 33:-*

† Refers to users who join the discussion due to being influenced by political inclination of entities (i) which have appeared in previously co-occurred articles or in the discussion of current article, but (ii) have never appeared in the main text of current article.

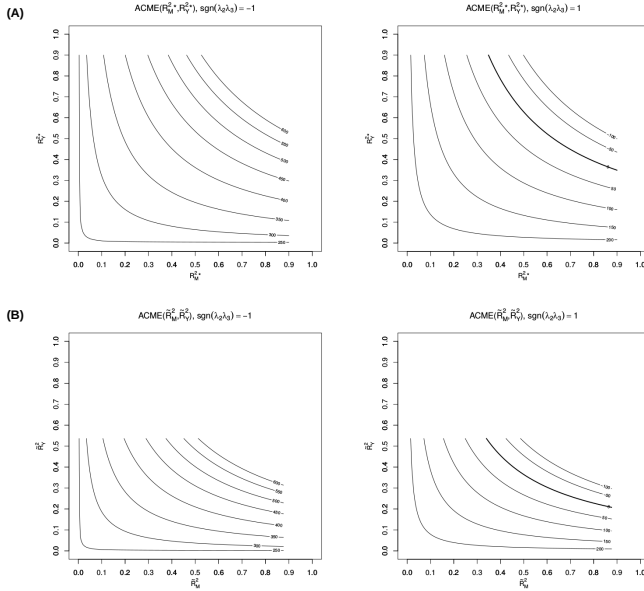
**Figure 24: SAMPLE SENSITIVITY TEST VISUALIZATION: PART 1**



This shows the variation of ACME of Model 2 in Table 32 for the indirect type micro attribute as a function of the sensitivity parameter  $\rho$ , the correlation between the error terms in the mediator and the outcome models. If small deviations from  $\rho = 0$  create huge variations in ACME, then our estimates might be sensitive to violation of sequential ignorability assumption. We see that this is not the case here. The solid line and gray band represent point estimates of ACME and their 90% confidence intervals, respectively. The estimated ACME is quite robust because it will turn zero only when  $\rho$  is 0.6. The current estimate of ACME, as shown in Model 2 of Table 32 for indirect type, assumes sequential ignorability assumption because the estimate has been computed with  $\rho = 0$ . What this figure shows is that our estimates (sign, confidence) can become invalid if  $\rho$  is 0.6.



**Figure 25: SAMPLE SENSITIVITY TEST VISUALIZATION: PART 2**



This shows the variation of ACME of Model 2 in Table 32 for the indirect type micro attribute as functions of  $(\tilde{R}_Y^2, \tilde{R}_M^2)$  and  $(R_Y^{2*}, R_M^{2*})$  in panels (A) and (B) respectively.  $\tilde{R}^2$  represents the proportion of *total* variance in the outcome ( $Y$ ) or the mediator ( $M$ ) variable that could be explained by an unobserved pretreatment confounder.  $R^{2*}$  represents the proportion of *unexplained* variance in the outcome ( $Y$ ) or the mediator ( $M$ ) variable that could be explained by an unobserved pretreatment confounder. The left and right panes represent negative and positive signs of the product of the coefficients of a potential unobserved confounder. The contour line of 0, in respective panels, correspond to pairs of values of  $(\tilde{R}_Y^2, \tilde{R}_M^2)$  and  $(R_Y^{2*}, R_M^{2*})$  for which ACME would be zero. We see that any unobserved confounder would have to explain a large chunk of  $\tilde{R}^2$  or  $R^{2*}$  for both outcome and mediator in order for ACME to be 0. The plots suggest that, under assumption of sequential ignorability, the estimates are quite robust to a potential unobserved pretreatment mediator-outcome confounding to a large extent.

# References

- [1] Abhishek Samantray and Massimo Riccaboni. Peer influence in large dynamic network: Quasi-experimental evidence from scratch. In Luca Maria Aiello, Chantal Cherifi, Hocine Cherifi, Renaud Lambiotte, Pietro Lió, and Luis M. Rocha, editors, *Complex Networks and Their Applications VII*, pages 300–313, Cham, 2019. Springer International Publishing.
- [2] Abhishek Samantray and Massimo Riccaboni. Peer influence of production and consumption behaviour in an online social network of collective learning. *Online Social Networks and Media*, 18: 100088, 2020. ISSN 2468-6964. doi: <https://doi.org/10.1016/j.osnem.2020.100088>. URL <http://www.sciencedirect.com/science/article/pii/S246869642030029X>.
- [3] Winter Mason and Duncan J. Watts. Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, 109(3):764–769, 2012. doi: 10.1073/pnas.1110069108.
- [4] Jan Feld and Ulf Zölitz. Understanding peer effects - on the nature, estimation and channels of peer effects. Technical Report ROA-RM-2016/1, Maastricht University, 2016.
- [5] Bruce Sacerdote. Peer effects with random assignment: Results for dorm-mates. *The Quarterly Journal of Economics*, 116(2):681–704, 2001.
- [6] Rabbany Reihaneh, Samira Elatia, Mansoureh Takaffoli, and Osmar R. Zaiane. *Educational Data Mining: Applications and Trends*, chapter Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective, pages 441–466. Springer International Publishing, 2014. doi: 10.1007/978-3-319-02738-8\_16.
- [7] Benjamin Mako Hill and Andrés Monroy-Hernández. A longitudinal dataset of five years of public activity in the scratch online community. *Scientific Data*, 4(170002), 2017. doi: 10.1038/sdata.2017.2.

- [8] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012. doi: 10.1038/nature11421.
- [9] Dean Eckles, René F. Kizilcec, and Eytan Bakshy. Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences*, 113(27):7316–7322, 2016. doi: 10.1073/pnas.1511201113.
- [10] Sinan Aral and Dylan Walker. Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science*, 60(6):1352–1370, 2014. doi: 10.1287/mnsc.2014.1936.
- [11] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014. doi: 10.1073/pnas.1320040111.
- [12] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010. doi: 10.1126/science.1185231.
- [13] Charles F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.
- [14] Tom A.B. Snijders, Gerhard G. van de Bunt, and Christian E.G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010.
- [15] Tom A.B. Snijders. Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application*, 4(1):343–63, 2017. doi: 10.1146/annurev-statistics-060116-054035.
- [16] Christian Steglich, Tom A. B. Snijders, and Michael Pearson. Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393, 2010.
- [17] Tom A.B. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 11(37):131–53, 2011. doi: 10.1146/annurev.soc.012809.102709.
- [18] Kevin Lewis, Marco Gonzalez, and Jason Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, 2012. doi: 10.1073/pnas.1109739109.
- [19] Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55, 2009.

- [20] Bryan S. Graham. Methods of identification in social networks. Technical Report 20414, NBER Working Paper, 2014.
- [21] Charles F. Manski. Identification problems in the social sciences. *Sociological Methodology*, 23(1):1–56, 1993.
- [22] Camila F. S. Campos, Shaun Hargreaves Heap, and Fernanda Leite Lopez de Leon. The political influence of peer groups: experimental evidence in the classroom. *Oxford Economic Papers*, 69(4):963–985, 2017. doi: 10.1093/oep/gpw065.
- [23] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334, 2010.
- [24] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009. doi: <https://doi.org/10.1073/pnas.0908800106>.
- [25] Wikipedia. Scratch. [https://en.wikipedia.org/wiki/Scratch\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/Scratch_(programming_language)), May 2018. [Last accessed 31-05-2018].
- [26] Scratch-Wiki. Activity feeds. [https://en.scratch-wiki.info/wiki/Activity\\_Feeds](https://en.scratch-wiki.info/wiki/Activity_Feeds), May 2018. [Last accessed 31-05-2018].
- [27] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2): 026126, 2003. doi: 10.1103/PhysRevE.67.026126.
- [28] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE*, 9(6): e98679, 2014. doi: 10.1371/journal.pone.0098679.
- [29] Luca Maria Aiello, Alain Barrat, Rossano Schifanella, C. Cattuto, Benjamin Markines, and Filippo Menczer. Friendship prediction and homophily in social media. *ACM Transaction on the Web*, 6(2):9:1–9:33, 2012. doi: 10.1145/2180861.2180866.
- [30] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, 2011. doi: 10.1177/0049124111404820.
- [31] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1): 415–444, 2001. doi: 10.1146/annurev.soc.27.1.415.

- [32] Maureen T. Hallinan and Richard A. Williams. Students' characteristics and the peer-influence process. *Sociology of Education*, 63(2):122–132, 1990. doi: 10.2307/2112858.
- [33] Robert Huckfeldt and John Sprague. Networks in context: The social flow of political information. *The American Political Science Review*, 81(4):1197–1216, 1987. doi: 10.2307/1962585.
- [34] Matthew D. Atkinson and Anthony Fowler. Social capital and voter turnout: Evidence from saint's day fiestas in mexico. *British Journal of Political Science*, 44(1):41–59, 2014. doi: 10.1017/S0007123412000713.
- [35] Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236, 2007. URL <http://gking.harvard.edu/files/abs/matchp-abs.shtml>.
- [36] Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart. Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 2011. URL <http://gking.harvard.edu/matchit/>.
- [37] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. mediation: R package for causal mediation analysis. *Journal of Statistical Software, Articles*, 59(5):1–38, 2014. doi: 10.18637/jss.v059.i05.
- [38] Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, 2010.
- [39] Jenny Davis. Prosuming identity: The production and consumption of transableism on transabled.org. *American Behavioral Scientist*, 56(4):596–617, 2012. doi: 10.1177/0002764211429361.
- [40] Andrea Forte, Judd Antin, Shaowen Bardzell, Leigh Honeywell, John Riedl, and Sarah Stierch. Some of all human knowledge: Gender and participation in peer production. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, CSCW '12, pages 33–36, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1051-2. doi: 10.1145/2141512.2141530. URL <http://doi.acm.org/10.1145/2141512.2141530>.
- [41] Julia Adams and Hannah Brückner. Wikipedia, sociology, and the promise and pitfalls of big data. *Big Data & Society*, 2(2):2053951715614332, 2015. doi: 10.1177/2053951715614332.

- [42] Csilla Rudas, Olivér Surányi, Taha Yasseri, and János Török. Understanding and coping with extremism in an online collaborative environment: A data-driven modeling. *PLOS ONE*, 12(3):1–16, 03 2017. doi: 10.1371/journal.pone.0173561.
- [43] Gerardo Iñiguez, János Török, Taha Yasseri, Kimmo Kaski, and János Kertész. Modeling social dynamics in a collaborative environment. *EPJ Data Science*, 3(1):7, Sep 2014. doi: 10.1140/epjds/s13688-014-0007-z.
- [44] Nabeel Gillani, Taha Yasseri, Rebecca Eynon, and Isis Hjorth. Structural limitations of learning in a crowd: communication vulnerability and information diffusion in moocs. *Scientific Reports*, 4(6447), 2015. doi: 10.1038/srep06447.
- [45] Lauren E. Sherman, Patricia M. Greenfield, Leanna M. Hernandez, and Mirella Dapretto. Peer influence via instagram: Effects on brain and behavior in adolescence and young adulthood. *Child Development*, 89(1):37–47, 2018. doi: 10.1111/cdev.12838.
- [46] Corey Phelps, Ralph Heidl, and Anu Wadhwa. Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, 38(4):1115–1166, 2012. doi: 10.1177/0149206311432640.
- [47] Todd Rogers and Avi Feller. Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting. *Psychological Science*, 27(3): 365–374, 2016. doi: 10.1177/0956797615623770.
- [48] Dan Davis, Ioana Jivet, René F. Kizilcec, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Follow the successful crowd: Raising mooc completion rates through social comparison at scale. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, LAK '17*, pages 454–463, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4870-6. doi: 10.1145/3027385.3027411.
- [49] Rhona Sharpe and Greg Benfield. The student experience of e-learning in higher education: A review of the literature. *Brookes eJournal of Learning and Teaching*, 1, 01 2005.
- [50] Aftab Dean and Andy Lima. Student experience of e-learning tools in he: An integrated learning framework. *European Journal of Social Science Education and Research*, 4(6):39–51, 2017. ISSN 2312-8429. doi: 10.26417/ejser.v11i2.p39-51.
- [51] Chia-Wen Tsai. Do students need teacher’s initiation in online collaborative learning? *Computers & Education*, 54(4):1137 – 1144, 2010. ISSN 0360-1315. doi: <https://doi.org/10.1016/j.compedu.2009.10.021>.

- [52] Arun Sundararajan, Foster Provost, Gal Oestreicher-Singer, and Sinan Aral. Research commentary - information in digital, economic, and social networks. *Information Systems Research*, 24(4):883–905, 2013. doi: 10.1287/isre.1120.0472.
- [53] Scratch-Wiki. Friend. <https://en.scratch-wiki.info/wiki/Friend>, May 2018. [Last accessed 31-05-2018].
- [54] Kevin Lewis, Marco Gonzalez, and Jason Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, 2012. doi: 10.1073/pnas.1109739109.
- [55] Peter J. Carrington, John Scott, and Stanley Wasserman, editors. *Structural Analysis in the Social Sciences*, page 329. Structural Analysis in the Social Sciences. Cambridge University Press, 2005. doi: 10.1017/CBO9780511811395.014.
- [56] Vikram Krishnamurthy, Omid Namvar Gharehshiran, and Maziyar Hamdi. Interactive sensing and decision making in social networks. *Foundations and Trends in Signal Processing*, 7(1-2):1–196, 2014. ISSN 1932-8346. doi: 10.1561/20000000048. URL <http://dx.doi.org/10.1561/20000000048>.
- [57] Dean Eckles. Identifying peer influence effects in observational social network data: An evaluation of propensity score methods. Technical report, Stanford University, 2010.
- [58] Dean Eckles and Eytan Bakshy. Bias and high-dimensional adjustment in observational studies of peer effects. Technical report, MIT, 2010.
- [59] Gary King and Richard Nielsen. Why propensity scores should not be used for matching. Technical report, Harvard University, 2016.
- [60] Bonnie Stewart. Open to influence: what counts as academic influence in scholarly networked twitter participation. *Learning, Media and Technology*, 40(3):287–309, 2015. doi: 10.1080/17439884.2015.1015547.
- [61] Ana Lucía Schmidt, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, 114(12):3035–3039, 2017. doi: 10.1073/pnas.1617052114.
- [62] Scott E. Carrell, Bruce I. Sacerdote, and James E. West. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855–882, 2013. doi: 10.3982/ECTA10168.

- [63] Abhishek Samantray and Paolo Pin. Credibility of climate change denial in social media. *Palgrave Communications*, 5:127, 2019. doi: 10.1057/s41599-019-0344-4. URL <https://doi.org/10.1057/s41599-019-0344-4>.
- [64] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Hao-han Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018. doi: 10.1073/pnas.1804840115.
- [65] Paul DiMaggio, John Evans, and Bethany Bryson. Have American’s social attitudes become more polarized? *American Journal of Sociology*, 102(3): 690–755, 1996.
- [66] Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in American public opinion. *American Journal of Sociology*, 114(2):408–446, 2008. doi: 10.1086/590649.
- [67] Markus Prior. Media and political polarization. *Annual Review of Political Science*, 16(1):101–127, 2013. doi: 10.1146/annurev-polisci-100711-135242.
- [68] Shanto Iyengar and Sean J. Westwood. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3):690–707, 2015. doi: 10.1111/ajps.12152.
- [69] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76 (3):405–431, 09 2012. doi: 10.1093/poq/nfs038.
- [70] Aaron M. McCright and Riley E. Dunlap. Anti-reflexivity. *Theory, Culture & Society*, 27(2-3):100–133, 2010. doi: 10.1177/0263276409356001.
- [71] Pew Research Center. Political polarization in the American public, June 2014.
- [72] Frances C. Moore, Nick Obradovich, Flavio Lehner, and Patrick Baylis. Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. *Proceedings of the National Academy of Sciences*, 116(11):4905–4910, 2019. doi: 10.1073/pnas.1816541116.
- [73] Aaron M. McCright, Riley E. Dunlap, and Chenyang Xiao. The impacts of temperature anomalies and political orientation on perceived winter warming. *Nature Climate Change*, 4:1077–1081, 2014. doi: 10.1038/nclimate2443.



- [74] Lawrence C. Hamilton. Education, politics and opinions about climate change evidence for interaction effects. *Climatic Change*, 104(2):231–242, Jan 2011. doi: 10.1007/s10584-010-9957-8.
- [75] Pablo Barberá. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1):76–91, 2015. doi: 10.1093/pan/mpu011.
- [76] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer. Political polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [77] Matthew O. Jackson and Dunia López-Pintado. Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(1):49–67, 2013. doi: 10.1017/nws.2012.7.
- [78] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. Users polarization on Facebook and Youtube. *PLoS ONE*, 11(8), 2016. doi: 10.1371/journal.pone.0159641.
- [79] Cameron E. Taylor, Alexander V. Mantzaris, and Ivan Garibay. Exploring how homophily and accessibility can facilitate polarization in social networks. *Information*, 9(325), 2018.
- [80] Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012. doi: 10.1093/qje/qjs021.
- [81] Jens Koed Madsen, Richard M Bailey, and Toby D. Pilditch. Large networks of rational agents form persistent echo chambers. *Scientific Reports*, 8(12391), 2018. doi: 10.1038/s41598-018-25558-7.
- [82] Daron Acemoglu, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4): 1201–1236, 2011. doi: 10.1093/restud/rdr004.
- [83] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. Spread of (mis)information in social networks. *Games and Economic Behavior*, 70:194–227, 2010.
- [84] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 U.S. Election: Divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, 2005. doi: 10.1145/1134271.1134277.

- [85] Lorien Jasny, Joseph Waggle, and Dana R. Fisher. An empirical examination of echo chambers in US climate policy networks. *Nature Climate Change*, 5:782–786, 2015. doi: 10.1038/nclimate2666.
- [86] R. Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, 14(2):265–285, 2009. doi: 10.1111/j.1083-6101.2009.01440.x.
- [87] Seth Flaxman, Sharad Goel, and Justin M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(51):298–320, 2016. doi: 10.1093/poq/nfw006.
- [88] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015. doi: 10.1126/science.aaa1160.
- [89] Hywel T.P. Williams, James R. McMurray, Tim Kurz, and F. Hugo Lambert. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32:126–138, 2015. doi: 10.1016/j.gloenvcha.2015.03.006.
- [90] Jennifer D. Greer. Evaluating the credibility of online information: A test of source and advertising influence. *Mass Communication and Society*, 6(1): 11–28, 2003. doi: 10.1207/S15327825MCS0601\3.
- [91] Spiro Kioussis. Public trust or mistrust? perceptions of media credibility in the information age. *Mass Communication and Society*, 4(4):381–403, 2001. doi: 10.1207/S15327825MCS0404\4.
- [92] Carl I. Hovland and Walter Weiss. The Influence of Source Credibility on Communication Effectiveness\*. *Public Opinion Quarterly*, 15(4):635–650, 01 1951. ISSN 0033-362X. doi: 10.1086/266350. URL <https://doi.org/10.1086/266350>.
- [93] Chanthika Pornpitakpan. The persuasiveness of source credibility: A critical review of five decades’ evidence. *Journal of Applied Social Psychology*, 34(2):243–281, 2004. doi: 10.1111/j.1559-1816.2004.tb02547.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.2004.tb02547.x>.
- [94] Martin Heesacker, Richard E. Petty, and John T. Cacioppo. Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality*, 51(4):653–666, 1983. doi: 10.1111/j.1467-6494.1983.tb00872.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.1983.tb00872.x>.

- [95] Hunt Allcott and Mathew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, 2017.
- [96] David Markham. The Dimensions of Source Credibility of Television Newscasters. *Journal of Communication*, 18(1):57–64, 02 2006. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1968.tb00055.x. URL <https://doi.org/10.1111/j.1460-2466.1968.tb00055.x>.
- [97] David Westerman, Patric R. Spence, and Brandon Van Der Heide. Social Media as Information Source: Recency of Updates and Credibility of Information\*. *Journal of Computer-Mediated Communication*, 19(2):171–183, 01 2014. ISSN 1083-6101. doi: 10.1111/jcc4.12041. URL <https://doi.org/10.1111/jcc4.12041>.
- [98] C.J. Hutto and Eric Gilbert. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media*, June 2014.
- [99] Parinaz Sobhani, Saif M. Mohammad, and Svetlana Kiritchenko. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 159–169, August 2016.
- [100] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3): 26:1–26:23, 2017. doi: 10.1145/3003433.
- [101] Haiko Lietz, Claudia Wagner, Arnim Bleier, and Markus Strohmaier. When politicians talk: Assessing online conversational practices of political parties on Twitter. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- [102] Pablo Aragón, Karolin Eva Kappler, Andreas Kaltenbrunner, David Laniado, and Yana Volkovich. Communication dynamics in Twitter during political campaigns: The case of the 2011 Spanish national election. *Policy & Internet*, 5(2):183–206, 2013. doi: 10.1002/1944-2866.POI327.
- [103] Sergio Currarini, Matthew O. Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4): 1003–1045, July 2009. doi: 10.3982/ECTA7528.
- [104] Yosh Halberstam and Brian Knight. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143:73–88, 2016. doi: 10.1016/j.jpubeco.2016.08.011.

- [105] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, 64(2):317–332, 2014. doi: 10.1111/jcom.12084.
- [106] Ema Kušen and Mark Strembeck. Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. *Online Social Networks and Media*, 5:37–50, 2018. doi: 10.1016/j.osnem.2017.12.002.
- [107] Yphtach Lelkes. Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410, 2016. doi: 10.1093/poq/nfw005.
- [108] Jonathan B. Freeman and Rick Dale. Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, 45(1):83–97, 2013.
- [109] Roland Pfister, Katharina A. Schwarz, Markus Janczyk, Rick Dale, and Jonathan B. Freeman. Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in Psychology*, 4:700, 2013.
- [110] Joan-María Esteban and Debraj Ray. On the measurement of polarization. *Econometrica*, 62(4):819–851, 1994. doi: 10.2307/2951734.
- [111] C.W.J. Granger and A.A. Weiss. Time series analysis of error-correcting models. In Samuel Karlin, Takeshi Amemiya, and Leo A. Goodman, editors, *Studies in Econometrics, Time Series, and Multivariate Statistics*, pages 255–278. Academic Press, 1983. ISBN 978-0-12-398750-1. doi: <https://doi.org/10.1016/B978-0-12-398750-1.50018-8>.
- [112] Robert F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987.
- [113] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.
- [114] Hiro Y. Toda and Taku Yamamoto. Statistical inference in vector autoregressions possibly integrated processes. *Journal of Econometrics*, 66:225–250, 1995.
- [115] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018. doi: 10.1126/science.aap9559.
- [116] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016. doi: 10.1073/pnas.1517441113.

- [117] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Homophily and polarization in the age of misinformation. *The European Physical Journal Special Topics*, 225(10):2047–2059, 2016. doi: 10.1140/epjst/e2015-50319-0.
- [118] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. Political rumor-ing on Twitter during the 2012 US presidential election: Rumor diffusion and correction. *New Media & Society*, 19(8):1214–1235, 2017. doi: 10.1177/1461444816634054.
- [119] Benjamin R. Warner. Segmenting the electorate: The effects of exposure to political extremism online. *Communication Studies*, 61(4):430–444, 2010. doi: 10.1080/10510974.2010.497069.
- [120] Elizabeth Dubois and Grant Blank. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5):729–745, 2018. doi: 10.1080/1369118X.2018.1428656.
- [121] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1395–1405, 2015. doi: 10.1145/2736277.2741637.
- [122] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. Computational fact checking from knowledge networks. *PLoS ONE*, 10(6), 2015. doi: 10.1371/journal.pone.0128193.
- [123] Endre Tvinnereim, Xiaozhi Liu, and Eric M. Jamelske. Public perceptions of air pollution and climate change: different manifestations, similar causes, and concerns. *Climatic Change*, 140(3-4):399–412, 2017. doi: 10.1007/s10584-016-1871-2.
- [124] Shanto Iyengar and Douglas S. Massey. Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences*, 116(16):7656–7661, 2019. doi: 10.1073/pnas.1805868115.
- [125] Ann E. Williams. Media evolution and public understanding of climate science. *Politics and the Life Sciences*, 30(2):20–30, 2016. doi: 10.2990/30\2\\_20.
- [126] Aaron M. McCright and Riley E. Dunlap. The politicization of climate change and polarization in the American public’s views of global warming, 2001-2010. *The Sociological Quarterly*, 52(2):155–194, 2011. doi: 10.1111/j.1533-8525.2011.01198.x.

- [127] Matthew A. Baum and Philip B.K. Potter. The relationships between mass media, public opinion, and foreign policy: Toward a theoretical synthesis. *Annual Review of Political Science*, 11(1):39–65, 2008. doi: 10.1146/annurev.polisci.11.060406.214132.
- [128] Gary King, Benjamin Schneer, and Ariel White. How the news media activate public expression and influence national agendas. *Science*, 358(6364): 776–780, 2017. doi: 10.1126/science.aao1100.
- [129] Wayne Wanta, Guy Golan, and Cheolhan Lee. Agenda setting and international news: Media influence on public perception. *Journalism and Mass Communication Quarterly*, 81(2):364–377, 2004. doi: 10.1177/107769900408100209.
- [130] Edgar Grande, Tobias Schwarzbözl, and Matthias Fatke. Politicizing immigration in western europe. *Journal of European Public Policy*, 26(10):1444–1463, 2018. doi: 10.1080/13501763.2018.1531909.
- [131] Swen Hutter and Edgar Grande. Politicizing europe in the national electoral arena: A comparative analysis of five west european countries, 1970-2010. *Journal of Common Market Studies*, 52(5):1002–1018, 2014. doi: 10.1111/jcms.12133.
- [132] Matteo Coronese, Francesco Lamperti, Klaus Keller, Francesca Chiaromonte, and Andrea Roventini. Evidence for sharp increase in the economic damages of extreme natural disasters. *Proceedings of the National Academy of Sciences*, 116(43):21450–21455, 2019. doi: 10.1073/pnas.1907826116.
- [133] D. Kahan. Why we are poles apart on climate change. *Nature*, 488, 2012. doi: 10.1038/488255a.
- [134] John Cook, Naomi Oreskes, Peter T Doran, William R L Anderegg, Bart Verheggen, Ed W Maibach, J Stuart Carlton, Stephan Lewandowsky, Andrew G Skuce, Sarah A Green, Dana Nuccitelli, Peter Jacobs, Mark Richardson, Bärbel Winkler, Rob Painting, and Ken Rice. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4), 2016. doi: 10.1088/1748-9326/11/4/048002.
- [135] Nick Pidgeon and Baruch Fischhoff. The role of social and decision sciences in communicating uncertain climate risks. *Nature Climate Change*, 1: 35–41, 2011. doi: 10.1038/nclimate1080.
- [136] Geoffrey Supran and Naomi Oreskes. Assessing Exxonmobil’s climate change communications (1977-2014). *Environmental Research Letters*, 12(8), 2017. doi: 10.1088/1748-9326/aa815f.

- [137] William R L Anderegg and Gregory R Goldsmith. Public interest in climate change over the past decade and the effects of the ‘climategate’ media event. *Environmental Research Letters*, 9(5), 2014. doi: 10.1088/1748-9326/9/5/054005.
- [138] Toby Bolsen and Matthew A. Shapiro. Strategic framing and persuasive messaging to influence climate change perceptions and decisions, 07 2017.
- [139] Maxwell T. Boykoff. From convergence to contention: United states mass media representations of anthropogenic climate change science. *Transactions of the Institute of British Geographers*, 32(4):477–489, 2007. doi: 10.1111/j.1475-5661.2007.00270.x.
- [140] Rodrigo Zamith, Juliet Pinto, and Maria Elena Villar. Constructing climate change in the americas: An analysis of news coverage in u.s. and south american newspapers. *Science Communication*, 35(3):334–357, 2012. doi: 10.1177/1075547012457470.
- [141] Aaron M. McCright and Riley E. Dunlap. The politicization of climate change and polarization in the american public’s views of global warming, 2001-2010. *The Sociological Quarterly*, 52(2):155–194, 2011. doi: 10.1111/j.1533-8525.2011.01198.x.
- [142] J Painter and N. Gavin. Climate skepticism in british newspapers, 2007-2011. *Environmental Communication*, 10(4):432–452, 2015. doi: 10.1080/17524032.2014.995193.
- [143] B. Tranter. Political divisions over climate change and environmental issues in Australia. *Environmental Politics*, 20(1):78–96, 2011. doi: 10.1080/09644016.2011.538167.
- [144] Aaron M. McCright, Riley E. Dunlap, and Sandra T. Marquart-Pyatt. Political ideology and views about climate change in the European Union. *Environmental Politics*, 25(2):338–358, 2015. doi: 10.1080/09644016.2015.1090371.
- [145] Craig Trumbo. Constructing climate change: claims and frames in US news coverage of an environmental issue. *Public Understanding of Science*, 5(3): 269–283, 1996. doi: 10.1088/0963-6625/5/3/006.
- [146] Peter Weingart, Anita Engels, and Petra Pansegrau. Risks of communication: discourses on climate change in science, politics, and the mass media. *Public Understanding of Science*, 9(3):261–283, 2000. doi: 10.1088/0963-6625/9/3/304.

- [147] Michael Brüggemann and Sven Engesser. Beyond false balance: How interpretive journalism shapes media coverage of climate change. *Global Environmental Change*, 42:58–67, 2017. doi: <https://doi.org/10.1016/j.gloenvcha.2016.11.004>.
- [148] Andrew S. Ross and Damian J. Rivers. Internet memes, media frames, and the conflicting logics of climate change discourse. *Environmental Communication*, 13(7):975–994, 2019. doi: 10.1080/17524032.2018.1560347.
- [149] Jason T. Carmichael, Robert J. Brulle, and Joanna K. Huxster. The great divide: understanding the role of media and other drivers of the partisan divide in public concern over climate change in the usa, 2001–2014. *Climatic Change*, 141(4):599–612, 2017. doi: 10.1007/s10584-017-1908-1.
- [150] Sandra T. Marquart-Pyatt, Rachael L. Shwom, Thomas Dietz, Riley E. Dunlap, Stan A. Kaplowitz, Aaron M. McCright, and Sammy Zahran. Understanding public opinion on climate change: A call for research. *Environment: Science and Policy for Sustainable Development*, 53(4):38–42, 2011. doi: 10.1080/00139157.2011.588555.
- [151] Toby Bolsen and James N. Druckman. Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, 21(3):389–402, 2018. doi: 10.1177/1368430217737855.
- [152] Matthew C. Nisbet. *Engaging in science policy controversies*, chapter 13. Routledge, 2014. doi: 10.4324/9780203483794.ch13.
- [153] Janis L. Dickinson, Rhiannon Crain, Steve Yalowitz, and Tammy M. Cherry. How framing climate change influences citizen scientists’ intentions to do something about it. *The Journal of Environmental Education*, 44(3):145–158, 2013. doi: 10.1080/00958964.2012.742032.
- [154] Augusta Isabella Alberici and Patrizia Milesi. Online discussion, politicized identity, and collective action. *Group Processes & Intergroup Relations*, 19(1):43–59, 2016. doi: 10.1177/1368430215581430.
- [155] Kathryn Harrison. The road not taken: Climate change policy in canada and the united states. *Global Environmental Politics*, 7(4):92–117, 2007. doi: 10.1162/glep.2007.7.4.92.
- [156] Arthur Lupia. Communicating science in politicized environments. *Proceedings of the National Academy of Sciences*, 110(Supplement 3):14048–14054, 2013. doi: 10.1073/pnas.1212726110.

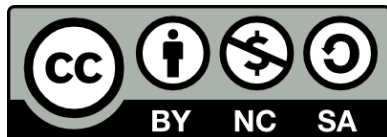


- [157] Aaron M. McCright and Riley E. Dunlap. Anti-reflexivity. the american conservative movement's success in undermining climate science and policy. *Theory, Culture & Society*, 27(2-3):100–133, 2010. doi: 10.1177/0263276409356001.
- [158] Dana R. Fisher, Philip Leifeld, and Yoko Iwaki. Mapping the ideological networks of american climate politics. *Climatic Change*, 116(3):523–545, 2013. doi: 10.1007/s10584-012-0512-7.
- [159] Douglas Guilbeault, Joshua Becker, and Damon Centola. Social learning and partisan bias in the interpretation of climate trends. *Proceedings of the National Academy of Sciences*, 115(39):9714–9719, 2018. doi: 10.1073/pnas.1722664115.
- [160] Amy E. Jasperson, Dhavan V. Shah, Mark Watts, Ronald J. Faber, and David P. Fan. Framing and the public agenda: Media effects on the importance of the federal budget deficit. *Political Communication*, 15(2):205–224, 1998. doi: 10.1080/10584609809342366.
- [161] Paul D'Angelo. *Framing: Media Frames*, pages 1–10. American Cancer Society, 2017. doi: 10.1002/9781118783764.wbieme0048.
- [162] Jörg Matthes and Matthias Kohring. The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58(2): 258–279, 2008. doi: 10.1111/j.1460-2466.2008.00384.x.
- [163] Clarissa C. David, Jenna Mae Atun, Erika Fille, and Christopher Monterola. Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures*, 5(4):329–351, 2011. doi: 10.1080/19312458.2011.624873.
- [164] Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1):89–106, 2016. doi: 10.1080/21670811.2015.1093271.
- [165] Dietram A. Scheufele and David Tewksbury. Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication*, 57(1):9–20, 2006. doi: 10.1111/j.0021-9916.2007.00326.x.
- [166] Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman, and Huan Liu. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2), 2018. doi: 10.1145/3204948.
- [167] Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research, November 2017.

- [168] Stefano M. Iacus, Gary King, and Giuseppe Porro. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1): 1–24, 2012. doi: 10.1093/pan/mpr013.
- [169] Christian Fong, Chad Hazlett, and Kosuke Imai. Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Annals of Applied Statistics*, 12(1):156–177, 03 2018. doi: 10.1214/17-AOAS1101.
- [170] Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1): 243–263, 2014. doi: 10.1111/rssb.12027.
- [171] Yeying Zhu, Donna L. Coffman, and Debashis Ghosh. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *Journal of Causal Inference*, 3(1):25–40, 2015. doi: <https://doi.org/10.1515/jci-2014-0022>.
- [172] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334, 2010.
- [173] Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25 (1):51–71, 2010.
- [174] Subal C. Kumbhakar and Ragnar Tveteras. Risk preferences, production risk and firm heterogeneity. *Scandinavian Journal of Economics*, 105(2):275–293, 2003.
- [175] Scratch-Wiki. Project copying. [https://wiki.scratch.mit.edu/wiki/Project\\_Copying](https://wiki.scratch.mit.edu/wiki/Project_Copying), May 2018. [Last accessed 31-05-2018].
- [176] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. URL <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008>.
- [177] Aaron Clauset, M. E. J. Newman, and Christopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6): 066111, December 2004. URL <https://link.aps.org/doi/10.1103/PhysRevE.70.066111>.
- [178] P. C. B. Phillips and S. Ouliaris. Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1):165–193, 1990.
- [179] Bernhard Pfaff. *Analysis of Integrated and Cointegrated Time Series with R*. Springer-Verlag New York, 2008. doi: 10.1007/978-0-387-75967-8.

- [180] Søren Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2):231–254, 1988. doi: 10.1016/0165-1889(88)90041-3.
- [181] Søren Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.
- [182] Erik Hjalmarsson and Pär Österholm. Testing for cointegration using the Johansen methodology when variables are near-integrated: size distortions and partial remedies. *Empirical Economics*, 39(1):51–76, 2010. doi: 10.1007/s00181-009-0294-6.
- [183] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- [184] C. A. Robertson and J. G. Fryer. Some descriptive properties of normal mixtures. *Scandinavian Actuarial Journal*, 3(4):137–146, 1969. doi: 10.1080/03461238.1969.10404590.
- [185] Benjamin Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, February 2010. doi: 10.1257/mic.2.1.112.





Unless otherwise expressly stated, all original material of whatever nature created by Abhishek Samantray and included in this thesis, is licensed under a Creative Commons Attribution Noncommercial Share Alike 3.0 Italy License.

Check on Creative Commons site:

<https://creativecommons.org/licenses/by-nc-sa/3.0/it/legalcode/>

<https://creativecommons.org/licenses/by-nc-sa/3.0/it/deed.en>

Ask the author about other uses.